METHOD OF MAKING PROTEIN ARRAYS

BACKGROUND OF THE INVENTION

This application was funded by DOE Grant No. DEFG02-87ER-60565. This application is a continuation-in-part of U.S. Patent Application No. 09/522,732, filed March 10, 2000, which in turn is a continuation-in-part of U.S. Patent Application No. 09/267,496, filed March 12, 1999, which in turn is a continuation-in-part of U.S. Patent Application No. 09/143,014, filed August 28, 1998. The application claims the benefit of U.S. Provisional Application No. 60/076,570, March 2, 1998 and U.S. Provisional Application No. 60/061,511, filed October 10, 1997.

1. Field of the Invention

Embodiments of the present invention relate to the production of protein arrays which may be used in low cost and/or high throughput methods for assessing protein structure or function in a research or diagnostic setting. More particularly, embodiments of the present invention are directed to the production of protein arrays by expressing nucleic acid arrays or by otherwise immobilizing proteins to nucleic acid arrays.

2. Description of Related Art

Arrays of nucleic acid molecules are of enormous utility in facilitating methods aimed at genomic characterization (such as polymorphism analysis and high-throughput sequencing techniques), screening of clinical patients or entire pedigrees for the risk of genetic disease, elucidation of protein/DNA- or protein/protein interactions or the assay of candidate pharmaceutical compounds for efficacy; however, such arrays are both labor-intensive and costly to produce by conventional methods. Highly ordered arrays of nucleic acid fragments are known in the art (Fodor et al., U.S. Patent No. 5,510,270; Lockhart et al., U.S. Patent No. 5,556,752). Chetverin and Kramer (WO 93/17126) are said to disclose a highly ordered array which may be amplified.

U.S. Patent No. 5,616,478 of Chetverin and Chetverina reportedly claims methods of nucleic acid amplification, in which pools of nucleic acid molecules are positioned on a

support matrix to which they are not covalently linked. Utermohlen (U.S. Patent No. 5,437,976) is said to disclose nucleic acid molecules randomly immobilized on a reusable matrix.

Different methods are known for attaching proteins to solid supports involving chemically linking such proteins to the solid support directly or via a linker molecule. See generally, Affinity Techniques, Enzyme Purification: Part B, Meth. Enz. 34 (ed. W.B. Jakoby and M. Wilchek, Acad. Press, N.Y. 1974) and Immobilized Biochemicals and Affinity Chromatography, Adv. Exp. Med. Biol. 42 (ed. R. Dunlap, Plenum Press, N.Y. 1974), incorporated by reference herein in their entireties for all purposes. US Patent No. 4,681,870 describes a method useful for covalently linking a protein to a surface of a silica matrix. US Patent No. 4,937,188 describes the use of RNA attached to a solid support where the RNA is reacted to a protein. US Patent No. 5,011,770 describes binding proteins that can be attached to a solid support. Roberts, R.W. and Szotak, J.W. Proc. Natl. Acad. Sci. (1997), 94(23):12297-12302 "RNA-Peptide Fusions for the In Vitro Selection of Peptides and Proteins" describes the in vitro translation of synthetic mRNAs to produce a covalent fusion between an mRNA and the peptide or protein that it encodes.

However, there is a need in the art to produce protein arrays whether by direct expression of the protein from a nucleic acid array or by binding proteins to a nucleic acid array, such as by hybridization or other means. On-demand expression of protein arrays for immediate use avoids protein stability problems that can be encountered with manufactured protein arrays that are required to be stored and/or shipped prior to use.

BRIEF SUMMARY OF THE INVENTION

Embodiments of the present invention are directed to methods of producing one or more arrays of proteins. Methods of the present invention include producing proteins from an array of nucleic acids. Alternatively, proteins may be immobilized to nucleic acid arrays to produce protein arrays. According to one embodiment of the present invention, a first nucleic acid array comprising nucleic acid molecules immobilized to a support is provided. The nucleic acid molecules can alternatively be random or ordered. Nucleic acid molecules may be deposited in an ordered fashion on discrete locations on the support by means well

known to those skilled in the art, such as by spotting or spraying a known concentration of nucleic acid molecules onto a known location of the support. Nucleic acid molecules may also be randomly deposited onto a support such as by simply spreading or spraying a mixture of nucleic acid molecules onto the surface of the support without predetermining the location for any particular nucleic acid. Each nucleic acid can then be amplified by polymerization reactions to produce a region of clones of the nucleic acid. Amplifying nucleic acids is particularly advantageous when a substantially pure amount of the nucleic acid is desired. According to the methods of the present invention, the nucleic acid molecules deposited onto the support can be amplified *in situ*. The nucleic acid molecules present on the support, whether amplified or not, are then expressed to produce proteins which are immobilized to the nucleic acid upon production or can be can be immobilized directly to the support. Alternatively, proteins can be bound to the nucleic acid molecules to produce protein arrays of the present invention. According to a certain aspect of the invention, arrays can include both nucleic acids and proteins or the nucleic acids can be removed from the array leaving the proteins.

According to an alternate embodiment of the present invention, a subset of the proteins that are produced can be transferred to an additional support where they are then immobilized. Accordingly, a method is provided whereby a first nucleic acid array comprising nucleic acid molecules immobilized to a support is provided. The nucleic acid molecules can alternatively be random or ordered. Optionally, the nucleic acid molecules are then amplified *in situ* to produce a colony of nucleic acid clone molecules as discussed above, such as by polymerization methods. A colony of nucleic acid molecules produced by polymerization methods is referred to herein as a "polony." The nucleic acid molecules on the support whether amplified or not are then expressed to produce proteins. A subset of the expressed proteins are then transferred to an additional support where they may then be immobilized.

An additional embodiment of the present invention is further provided where nucleic acids whether deposited or amplified can be transferred to an additional support where they are then optionally amplified and then expressed to produce proteins. Accordingly, a first nucleic acid array comprising nucleic acid molecules immobilized to a support is provided.

The nucleic acid molecules can alternatively be random or ordered. Optionally, the nucleic acid molecules are then amplified *in situ*. A subset of the nucleic acid molecules are then transferred to an additional support. The nucleic acid molecules are then immobilized, optionally amplified, and expressed to produce proteins. The proteins can then alternatively be immobilized or a subset can be transferred to yet another support where they can be immobilized.

According to certain aspects of the present invention, nucleic acids or expressed proteins can either be immobilized or transferred to an additional or subsequent support. Supports within the scope of the present invention can be either solid supports, such as flat slides, chips, beads or fibers, or semi-solid supports, such as gel matrices. Nucleic acids or proteins can be transferred by, among other methods, directly contacting supports.

Other features and advantages of certain embodiments of the present invention will become more fully apparent from the following description taken in conjunction with the accompanying figures and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

In the course of the detailed description of certain preferred embodiments to follow, reference will be made to the attached figures, in which.

Figure 1 shows the results six cycles of nucleotide addition and detection in polyacrylamide gel matrix fluorescent sequencing reactions on two different template nucleic acid samples. The top panel shows a fluorescent scan of the array after addition of fluorescently labeled dCTP, and the bottom panel shows schematics of sequencing template samples 1 and 2 with expected extension products.

Figure 2 shows the result of the addition of fluorescently labeled TTP in the eighth cycle of addition, detection, and cleavage in polyacrylamide gel matrix fluorescent sequencing reactions when the next correct nucleotide was an A. The top panel shows a fluorescent scan, and the bottom panel shows schematics of the expected extension products for sequencing template samples 1 and 2.

Figure 3 shows the result of the addition of fluorescently labeled dCTP in the tenth cycle of addition, detection and cleavage in polyacrylamide gel matrix fluorescent sequencing reactions of template samples 1 and 2. The panels are arranged as in Figure 2.

Figure 4 shows the result of the addition of fluorescently labeled TTP in the twelfth cycle of addition, detection and cleavage in polyacrylamide gel matrix fluorescent sequencing reactions of template samples 1 and 2. The panels are arranged as in Figure 2.

Figure 5 is a schematic drawing of a disulfide-bonded cleavable nucleotide fluorophore complex useful in the methods of the invention.

Figure 6 shows the results of experiments establishing the function of cleavable linkers in polyacrylamide gel matrix fluorescent sequencing reactions. The top panels show fluorescent scans of primer extension reactions, on two separate sequencing templates, in polyacrylamide spots using nucleotides with non-cleavably (Cy5-dCTP) and cleavably (Cy5-SS-dCTP) linked fluorescent label, before and after cleavage with dithiothreitol (DTT). The bottom panel shows schematics of sequencing templates 1 and 2 with the expected extension products.

Figure 7 is a schematic drawing of a nucleic acid template useful in making arrays according to the invention. Two constant regions flank a region of variable sequence.

Figure 8 shows the amplification of array features within a gel matrix. Figure 8A shows amplified arrays made using various amounts of starting template nucleic acid. Figure 8B shows the linear relationship between the amount of starting template nucleic acid and the number of amplified array features. Figure 8C shows an agarose gel containing PCR amplification products from a picked and re-amplified array feature.

Figure 9 shows the results of experiments examining the relationship of amplified feature size to template length and gel concentration. Figure 9A shows a plot of the radius of array features versus the log of the template length. Figure 9B shows array features created from a 1009 base pair template in a 15% polyacrylamide matrix.

Figure 10 shows a replica of a nucleic acid array made in a polyacrylamide gel matrix according to the methods of the invention. Figure 10A shows the original array, and Figure 10B shows a replica of the array of Figure 10A.

DETAILED DESCRIPTION OF CERTAIN PREFERRED EMBODIMENTS

The principles of the present invention may be applied with particular advantage to prepare protein arrays which can be used in methods for assessing protein structure or function in certain high throughput methods. According to the teachings of the present invention, an array of nucleic acids, i.e. whole cDNAs, partial cDNAs, modified cDNAs, synthetic nucleic acid sequences, naturally occurring nucleic acid sequences, chromosomes, RNA, mRNA, naturally occurring RNA, synthetic RNA, etc., is prepared according to various methods described in copending parent application U.S.S.N. 09/522,732 hereby incorporated by reference in its entirety for all purposes. Individual nucleic acids can be immobilized to the array substrate through covalent bonding, intermediate linker molecules, steric hindrance for example by means of a gel, hybridization or any combination thereof. Arrays of nucleic acids can also be produced according to methods well known in the art, such as Fodor et al., U.S. Patent No. 5,510,270, Lockhart et al., U.S. patent No. 5,556,752 and Chetverin and Kramer WO 93/17126 each of which are hereby incorporated by reference in their entireties for all purposes.

Once produced, the nucleic acid arrays can be amplified according to various methods described in copending parent application U.S.S.N. 09/522,732. The nucleic acid arrays can also be characterized for features, such as the sequence, function or chemical activity of the nucleic acids, on the arrays by any number of the methods described in parent application U.S.S.N. 09/522,732.

The nucleic acid arrays can be replicated according to various methods described in copending parent application U.S.S.N. 09/522,732. However, it is to be understood that methods of making protein arrays described herein need not incorporate any method step which replicates a nucleic acid or protein array.

Once the nucleic acid array has been produced, the nucleic acid array can then be used to produce proteins under appropriate conditions. Methods used to produce proteins from nucleic acids in vitro are known in the art and include Roberts, R.W. and Szostak, J.W., Proc. Natl. Acad. Sci. (1997), 94(23):12297-12302 "RNA-Peptide Fusions for the In Vitro Selection of Peptides and Proteins," Hanes et al., Nat. Biotechnol. (2000) 18:1287-92 "Picomolar Affinity Antibodies from a Fully Synthetic Naive Library Selected and Evolved

by Ribosome Display," and Mattheakis et al., Proc. Natl. Acad. Sci. (1994), 91:9022-9026 "An In Vitro Polysome Display System for Identifying Ligands from Very Large Peptide Libraries," each of which is hereby incorporated by reference in its entirety for all purposes. The proteins, in turn, are immobilized to the mRNA or they can then be immobilized to a support using methods known to those skilled in the art. Features of the protein arrays can be determined using methods known to those skilled in the art.

According to an additional embodiment of the present invention, proteins can be immobilized to nucleic acid arrays to produce protein arrays. Methods of immobilizing proteins to nucleic acids include, for example, hybridization of protein-mRNA conjugates to nucleic acids.

GLOSSARY OF TERMS

As used herein in reference to nucleic acid or protein arrays, the term "plurality" is defined as designating two or more such arrays, wherein a first (or "template") array plus a second array made from it comprise a plurality. When such a plurality comprises more than two arrays, arrays beyond the second array may be produced using either the first array or any copy of it as a template. As used herein, the terms "randomly-patterned" or "random" refer to a non-ordered, non-Cartesian distribution (in other words, not arranged at pre-determined points along the x- and y axes of a grid or at defined 'clock positions', degrees or radii from the center of a radial pattern) of nucleic acid or protein molecules over a support, that is not achieved through an intentional design (or program by which such a design may be achieved) or by placement of individual nucleic acid or protein features. Such a "randomly-patterned" or "random" array of nucleic acids or proteins may be achieved by dropping, spraying, plating or spreading a solution, emulsion, aerosol, vapor or dry preparation comprising a pool of nucleic acid molecules or proteins onto a support and allowing the nucleic acid molecules or proteins to settle onto the support without intervention in any manner to direct them to specific sites thereon. Protein arrays may be randomly patterned or random by virtue of the randomly patterned or random nature of the nucleic acid arrays used to generate the protein arrays.

As used herein, the term "ordered" when used to describe an array refers to those arrays achieved through an intentional design (or program by which such a design may be achieved) or by placement of individual nucleic acid or protein features at known locations on the array substrate. Methods of producing ordered arrays are well known and include such methods as spotting or spraying a known amount of a known nucleic acid or protein at a known location on a substrate. Devices used to create ordered arrays include those which are capable of depositing nucleic acids or proteins according to coordinates on a Cartesian grid. Methods are also known where ordered arrays are produced by surface modification of the support to create discrete regions on the support where nucleic acids or proteins are deposited and retained. Discrete regions can be created by laser or mechanical etching, photogelation, photomasking, deposition of hydrophobic and hydrophilic layers, and other methods known to those skilled in the art to confine deposited nucleic acids or proteins to discrete regions.

As used herein, the terms "immobilized" or "affixed" refer to an association between a nucleic acid or protein and a substrate characterized by covalent bonding, intermediate linker molecules, steric hindrance, hybridization or any combination thereof. For example, a nucleic acid or protein can be immobilized to a support by covalent bonding directly to the surface of the support which may or may not be modified to enhance such covalent bonding. Also, the nucleic acid or protein can be immobilized to the support by use of a linker molecule between the nucleic acid or protein and the support. Nucleic acids and proteins can further be immobilized on the support by steric hindrance within a polymerized gel or by covalent bonding within a polymerized gel. Nucleic acids and proteins can also be immobilized on a support through hybridization between the nucleic acid or protein and a molecule, such as a nucleic acid or protein, that is immobilized on the support. Affixing or immobilizing nucleic acid molecules to a support can be performed using a covalent linker that is selected from the group that includes oxidized 3-methyl uridine, an acrylyl group and hexaethylene glycol. Additionally, acrydite oligonucleotide primers may be covalently fixed within a polyacrylamide gel. It is also contemplated that affixing of nucleic acid molecules to the support is performed *via* hybridization of the members of the pool to nucleic acid molecules that are covalently bound to the support.

As used herein, the term "array" refers to a heterogeneous or homogenous pool of nucleic acid or protein molecules that is distributed over a support matrix; preferably, these molecules whether differing in sequence or not are spaced at a distance from one another sufficient to permit the identification of discrete features of the array.

As used herein, the term "heterogeneous" is defined to refer to a population or collection of nucleic acid or protein molecules that comprises a plurality of different sequences; it is contemplated that a heterogeneous pool of nucleic acid molecules results from a preparation of RNA or DNA from a cell which may be unfractionated or partially-fractionated.

An "unfractionated" nucleic acid preparation is defined as that which has not undergone the selective removal of any sequences present in the complement of RNA or DNA, as the case may be, of the biological sample from which it was prepared. A nucleic acid preparation in which the average molecular weight has been lowered by cleaving the component nucleic acid molecules, but which still retains all sequences, is still "unfractionated" according to this definition, as it retains the diversity of sequences present in the biological sample from which it was prepared.

A "partially-fractionated" nucleic acid preparation may have undergone qualitative size-selection. In this case, uncleaved sequences, such as whole chromosomes or RNA molecules, are selectively retained or removed based upon size. In addition, a "partially-fractionated" preparation may comprise molecules that have undergone selection through hybridization to a sequence of interest; alternatively, a "partially-fractionated" preparation may have had undesirable sequences removed through hybridization. It is contemplated that a "partially-fractionated" pool of nucleic acid molecules will not comprise a single sequence that has been enriched after extraction from the biological sample to the point at which it is pure, or substantially pure.

In this context, "substantially pure" as it relates to a nucleic acid sequence refers to a single nucleic acid sequence that is represented by a majority of nucleic acid molecules in a discrete region or in a pool. Substantially pure nucleic acid sequences can be obtained by *in vitro* polymerization. However, if a given sequence is heavily represented in the biological

sample, a preparation containing it is not excluded from being described as substantially pure according to the invention.

As used herein, the term "biological sample" refers to a whole organism or a subset of its tissues, cells or component parts (e.g., fluids). "Biological sample" further refers to a homogenate, lysate or extract prepared from a whole organism or a subset of its tissues, cells or component parts, or a fraction or portion thereof. Lastly, "biological sample" refers to a medium, such as a nutrient broth or gel in which an organism has been propagated, which contains cellular components, such as nucleic acid molecules.

As used herein, the term "organism" refers to all cellular life-forms, such as prokaryotes and eukaryotes, as well as non-cellular, nucleic acid-containing entities, such as bacteriophage and viruses.

As used herein, the term "feature" refers to each nucleic acid or protein sequence occupying a discrete physical location on the array; if a given sequence is represented at more than one such site, each site is classified as a feature. In this context, the term "nucleic acid sequence" may refer either to a single nucleic acid molecule, whether double or single-stranded, to a "clone" of amplified copies of a nucleic acid molecule present at the same physical location on the array (*i.e.*, a "polony") or to a replica, on a separate support, of such a clone.

As used herein, the term "protein" is used according to its ordinary meaning to refer to a compound whether natural or synthetic that is composed of a two or more amino acids joined by peptide linkages. The term "protein" includes peptides and/or oligopeptides and also includes modified and derivatized species..

As used herein, the term "amplifying" refers to production of copies of a nucleic acid molecule of the array *via* repeated rounds of primed enzymatic synthesis; *"in situ* amplification" indicates that such amplifying takes place with the template nucleic acid molecule positioned on a support according to the invention, rather than in solution.

As used herein, the term "support" refers to a substrate upon which nucleic acid or protein molecules are immobilized. The support may be rigid, solid or semi-solid. Virtually any solid substrate can be employed in the method of the invention. The substrate can be biological, nonbiological, organic, inorganic, or a combination of any of these materials,

existing as particles, strands, precipitates, gels, sheets, tubing, spheres, containers, capillaries, pads, slices, films, chips, plates, slides, etc. The substrate can have any convenient shape, such as a disc, square, sphere, circle, etc. The substrate and the surface of the substrate preferably form a rigid support on which to carry out the reactions described herein. Other substrate materials will be readily apparent to those of skill in the art upon review of this disclosure. In a preferred embodiment the substrate is flat glass or single-crystal silicon with surface features of less than 10 angstroms. Surfaces on the solid substrate will usually, though not always, be composed of the same material as the substrate. Thus, the surface may be composed of any of a wide variety of materials, for example, polymers, plastics, resins, polysaccharides, silica or silica-based materials, carbon, metals, inorganic glasses, membranes, etc., provided only that caged thiols can be attached to the surface of the substrate. Preferably, the surface will contain reactive groups, which could be carboxyl, amino, hydroxyl, or the like. Most preferably, the surface will be optically transparent and will have surface Si--OH functionalities, such as are found on silica surfaces. The surface of the support can be modified according to methods known to those skilled in the art to promote immobilization of nucleic acids or proteins thereon.

Supports can be essentially non-compressible and lacking pores containing liquid. A rigid or solid support can further be thin and thermally conductive, such that changes in thermal energy characteristic of PCR thermal cycling are conducted through the support to permit amplification of PCR template molecules arrayed on its surface.

As used herein, the term "semi-solid" refers to a compressible matrix with both a solid and a liquid component, wherein the liquid occupies pores, spaces or other interstices between the solid matrix elements. Semi-solid supports within the teachings of the present invention include starches, polyacrylamide, cellulose, polyamide (nylon) and cross-linked agarose, –dextran and –polyethylene glycol. Semi-solid supports can be combined with solid supports such as a glass slide combined with a polyacrylamide material.

As used herein in reference to the physical placement of nucleic acid or protein molecules or features and/or their orientation relative to one another on an array of the invention, the terms "correspond" or "corresponding" refer to a molecule occupying a position on a second array that is either identical to- or a mirror image of the position of a

molecule from which it was amplified on a first array which served as a template for the production of the second array, or vice versa, such that the arrangement of features of the array relative to one another is conserved between arrays of a plurality.

As implied by the above statement, a first and second array of a plurality of nucleic acid or protein arrays according to the invention may be of either like or opposite chirality, that is, the patterning of the nucleic acid or protein arrays may be either identical or mirror-imaged.

As used herein, the term "replica" refers to any nucleic acid or protein array that is produced by a printing process according to the invention using as a template a first randomly-patterned immobilized nucleic acid or protein array.

As used herein, the term "spot" as applied to a component of a microarray refers to a discrete area of a surface containing a substance deposited by mechanical or other means.

As used herein, "excluded volume" refers to the volume of space occupied by a particular molecule to the exclusion of other such molecules.

As used herein, "excess of nucleic acid molecules" refers to an amount of nucleic acid molecules greater than the amount of entities to which such nucleic acid molecules may bind. An excess may comprise as few as one molecule more than the number of binding entities, to twice the number of binding entities, up to 10 times, 100 times, 1000 times the number of binding entities or more.

As used herein, "signal amplification method" refers to any method by which the detection of a nucleic acid or protein is accomplished.

As used herein, a "nucleic acid capture ligand" or "nucleic acid capture activity" refers to any substance which binds nucleic acid molecules, either specifically or non-specifically, or which binds an affinity tag attached to a nucleic acid molecule in such a way as to immobilize the nucleic acid molecule to a support bearing the capture ligand.

As used herein, "replica-destructive" refers to methods of signal amplification which render an array or replica of an array non-reusable.

As used herein, the term "non-reusable," in reference to an array or replica of an array, indicates that, due to the nature of detection methods employed, the array cannot be

replicated nor used for subsequent detection methods after the first detection method is performed.

As used herein, the term "essentially distinct" as applied to features of an array refers to the situation where 90% or more of the features of an array are not in contact with other features on the same array.

As used herein, the term "preserved" as applied to the resolution of nucleic acid or protein features on an array means that the features remain essentially distinct after a given process has been performed.

As used herein, the term "distinguishable" as applied to a label, refers to a labeling moiety which can be detected when among other labeling moieties.

As used herein, the term "spectrally distinguishable" or "spectrally resolvable" as applied to a label, refers to a labeling moiety which can be detected by its characteristic fluorescent excitation or emission spectra, one or both of such spectra distinguishing said moiety from other moieties used separately or simultaneously in the particular method.

As used herein, the term "chain-terminating analog" refers to any nucleotide analog which, once incorporated onto the 3' end of a nucleic acid molecule, cannot serve as a substrate for further addition of nucleotides to that nucleic acid molecule.

As used herein, the term "type IIS" refers to a restriction enzyme that cuts at a site remote from its recognition sequence. Such enzymes are known to cut at a distances from their recognition sites ranging from 0 to 20 base pairs.

As used herein, the term "synthetic oligonucleotide" refers to a short (10 to 1,000 nucleotides in length), double- or single-stranded nucleic acid molecule that is chemically synthesized or is the product of a biological system such as a product of primed or unprimed enzymatic synthesis.

As used herein, the term "template DNA" refers to a plurality of DNA molecules used as the starting material or template for manufacture of a nucleic acid array such as a polyacrylamide-immobilized nucleic acid array.

As used herein, the term "template nucleic acids" refers to a plurality of nucleic acid molecules used as the starting material or template for manufacture of a nucleic acid array.

As used herein, the term "amplification primer" refers to an oligonucleotide that may be used as a primer for amplification reactions. The term "PCR primer" refers to an oligonucleotide that may be used as a primer for the polymerase chain reaction. A PCR primer is preferably, but not necessarily, synthetic, and will generally be approximately 10 to 100 nucleotides in length.

As used herein, the term "Acrydite modified" in reference to an oligonucleotide means that the oligonucleotide has an Acrydite phosphoramidite group attached to the 5' end of the molecule.

As used herein, the term "thermostable, template-dependent DNA polymerase" refers to an enzyme capable of conducting primed enzymatic synthesis following incubation at a temperature, greater than 65°C and less than or equal to approximately 100°C, and for a time, ranging from about 15 seconds to about 5 minutes, that is sufficient to denature essentially all double stranded DNA molecules in a given population. The term "isothermal" when used to describe certain methods herein means that the method does not require thermal cycling.

As used herein, the term "binding sites" when used in reference to a nucleic acid molecule, means sequences that hybridize under selected PCR annealing conditions with a selected PCR primer. Binding sites for PCR primers are generally used in pairs situated on either side of a sequence to be amplified, with each member of the pair preferably comprising a sequence from the other member of the pair.

As used herein, the term "variable sequence" refers to a sequence in a population of nucleic acid molecules that varies between different members of the population. Generally, as used herein, a variable sequence is flanked on either side by sequences that are shared or constant among all members of that population.

The following examples are set forth as representative of the present invention. These examples are not to be construed as limiting the scope of the invention as these and other equivalent embodiments will become apparent in view of the present disclosures, figures and accompanying claims.

EXAMPLE 1

Producing a Protein Array According to the Invention

A. Producing a Nucleic Acid Array for Use in the Invention

Step 1. Production of a Nucleic Acid Pool with Which to Construct an Array of Proteins/Peptides

A pool or library of n-mers (n= 20 to 9000) is made by any of several methods. The pool is either amplified (e.g., by PCR) or left unamplified. A suitable *in vitro* amplification "vector," for example, flanking PCR primer sequences or an in vivo plasmid, phage or viral vector from which amplified molecules are excised prior to use, is used. If necessary, random shearing or enzymatic cleavage of large nucleic acid molecules is used to generate the pools if the nucleic acid molecules are amplified, cleavage is performed either before or after amplification. Alternatively, a nucleic acid sample is random primed, for example with tagged 3' terminal hexamers followed by electrophoretic size-selection. The nucleic acid is selected from genomic, synthetic or cDNA sequences (Power, 1996, J. Hosp. Infect., 34: 247-265; Welsh, et al., 1995, Mutation Res., 338: 215-229). The copied or unamplified nucleic acid fragments resulting from any of the above procedures are, if desired, fractionated by size or affinity by a variety of methods including electrophoresis, sedimentation, and chromatography (possibly including elaborate, expensive procedures or limited-quantity resources since the subsequent inexpensive replication methods can justify such investment of effort).

Pools of nucleic acid molecules are, at this stage, applied directly to the support medium (see Step 2, below). Alternatively, they are cloned into nucleic acid vectors. For example, pools composed of fragments with inherent polarity, such as cDNA molecules, are directionally cloned into nucleic acid vectors that comprise, at the cloning site, oligonucleotide linkers that provide asymmetric flanking sequences to the fragments. Upon their subsequent removal *via* restriction with enzymes that cleave the vector outside both the cloned fragment and linker sequences, molecules with defined (and different) sequences at their two ends are generated. By denaturing these molecules and spreading them onto a semi-solid support to which is covalently bound oligonucleotides that are complementary to

one preferred flanking linker, the orientation of each molecule in the array is determined relative to the surface of the support. Such a polar array is of use for *in vitro* transcription/translation of the array or any purpose for which directional uniformity is preferred.

In addition to the attachment of linker sequences to the molecules of the pool for use in directional attachment to the support, a restriction site or regulatory element (such as a promoter element, cap site or translational termination signal), is, if desired, joined with the members of the pool. The use of fragments with termini engineered to comprise useful restriction sites is described below in Example 4.

Alternatively, the following methods can be used to create a library of full-length cDNAs bound to a substrate which can be used to make full-length proteins. Characterization of the nucleic acid sequence can then be accomplished as desired using, for example, FISSEQ. According to this method, a cDNA or genomic library of full length open reading frames (ORF) with a predetermined 3' junction can be enriched.

A full length transcript is obtained by using calf intestinal alkaline phosphatase followed by tobacco nucleotide acid pyro-phosphatase (TAP) with an RNA ligase reaction that links decapped mRNA to a specific oligoribonucleotide (38-mer). This 38-mer would contain a T7-promoter. Reverse transcriptase is then used to extend full length cDNA using one or more of the following three primers with 3' ends complementary to the three possible stop codons TAA, TGA, TAG):

5' GTGCAGNNNNNNNNNNNNTTA

5' GTGCAGNNNNNNNNNNNNCTA

and/or

5' GTGCAGNNNNNNNNNNNNTCA

RnaseH then prime the second strand cDNA with the 38 mer complement 3' overhang at one end of a double stranded fragment which contains the NEB IMPACT (Intein Mediated Purification with an Affinity Chitin-binding Tag). At the other end of the fragment is another BsgI site next to first intein codon (TGC = cys). The fragment can then be cleaved with BsgI, which creates two base 3' overhangs at the stop codon (TAR or TGA) and cys codon:

5' ATGNNN...NNNTA CTGCTTTGCCAAGGGTACCAATG

TACNNN...NNN ATGACGAAACGGTTCCCATGGTTAC 5'

and/or

5' ATGNNN...NNNTG CTGCTTTGCCAAGGGTACCAATG

TACNNN...NNN CTGACGAAACGGTTCCCATGGTTAC 5'

The cDNA is then circularized with T4 DNA ligase, transcribed with T7 RNAP, and translated according to the methods described in Roberts, R.W. and Szotak, J.W. Proc. Natl. Acad. Sci. (1997), 94(23):12297-12302 "RNA-Peptide Fusions for the In Vitro Selection of Peptides and Proteins." The mRNA protein hybrid is then selected on a chitin column by eluting with DTT (which cleaves at the intein cys).

The selection may then be repeated or the in vitro selected RNAs can be cloned into expression vectors or other suitable expression cassettes. A mutagenic primer and/or BsgI can be used to modify the junction codon (originally a stop codon, now a tyr or cys codon).

Step 2. Transfer of Nucleic Acid Pool onto a Support Medium

The nucleic acid pool is diluted ("plated") out onto a semi-solid medium (such as a polyacrylamide gel) on a solid surface such as a glass slide such that amplifiable molecules are 0.1 to 100 micrometers apart. Sufficient spacing is maintained that features of the array do not contaminate one another during repeated rounds of amplification and replication. It is estimated that a molecule that is immobilized at one end can, at most, diffuse the distance of a single molecule length during each round of replication. Obviously, arrays of shorter molecules are plated at higher density than those comprising long molecules.

Immobilizing media that are of use according to the invention are physically stable and chemically inert under the conditions required for nucleic acid molecule deposition, amplification and the subsequent replication of the array. A useful support matrix withstands the rapid changes in- and extremes of temperature required for PCR and retains structural integrity under stress during the replica printing process. The support material permits enzymatic nucleic acid synthesis; if it is unknown whether a given substance will do so, it is tested empirically prior to any attempt at production of a set of arrays according to the invention. The support structure comprises a semi-solid (i.e. gelatinous) lattice or matrix, wherein the interstices or pores between lattice or matrix elements are filled with an aqueous

or other liquid medium; typical pore (or 'sieve') sizes are in the range of 100 μm to 5 nm. Larger spaces between matrix elements are within tolerance limits, but the potential for diffusion of amplified products prior to their immobilization is increased. The semi-solid support is compressible, so that full surface-to-surface contact, essentially sufficient to form a seal between two supports, although that is not the object, may be achieved during replica printing. The support is prepared such that it is planar, or effectively so, for the purposes of printing; for example, an effectively planar support might be cylindrical, such that the nucleic acids of the array are distributed over its outer surface in order to contact other supports, which are either planar or cylindrical, by rolling one over the other. Lastly, a support materials of use according to the invention permits immobilizing (covalent linking) of nucleic acid features of an array to it by means enumerated below. Materials that satisfy these requirements comprise both organic and inorganic substances, and include, but are not limited to, polyacrylamide, cellulose and polyamide (nylon), as well as cross-linked agarose, dextran or polyethylene glycol.

Of the support media upon which the members of the pool of nucleic acid molecules may be anchored, one that is particularly preferred is a thin, polyacylamide gel on a glass support, such as a plate, slide or chip. A polyacrylamide sheet of this type is synthesized as follows: Acrylamide and bis-acrylamide are mixed in a ratio that is designed to yield the degree of crosslinking between individual polymer strands (for example, a ratio of 38:2 is typical of sequencing gels) that results in the desired pore size when the overall percentage of the mixture used in the gel is adjusted to give the polyacrylamide sheet its required tensile properties. Polyacrylamide gel casting methods are well known in the art (see Sambrook et al., 1989, Molecular Cloning. A Laboratory Manual., 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), and one of skill has no difficulty in making such adjustments.

The gel sheet is cast between two rigid surfaces, at least one of which is the glass to which it will remain attached after removal of the other. The casting surface that is to be removed after polymerization is complete is coated with a lubricant that will not inhibit gel polymerization; for this purpose, silane is commonly employed. A layer of silane is spread upon the surface under a fume hood and allowed to stand until nearly dry. Excess silane is

then removed (wiped or, in the case of small objects, rinsed extensively) with ethanol. The glass surface which will remain in association with the gel sheet is treated with γ-methacryloxypropyltrimethoxysilane (Cat. No. M6514, Sigma; St. Louis, MO), often referred to as 'crosslink silane,' prior to casting. The glass surface that will contact the gel is triply-coated with this agent. Each treatment of an area equal to 1200 cm^2 requires 125 µl of crosslink silane in 25 ml of ethanol. Immediately before this solution is spread over the glass surface, it is combined with a mixture of 750 µl water and 75 µl glacial acetic acid and shaken vigorously. The ethanol solvent is allowed to evaporate between coatings (about 5 minutes under a fume hood) and, after the last coat has dried, excess crosslink silane is removed as completely as possible *via* extensive ethanol washes in order to prevent 'sandwiching' of the other support plate onto the gel. The plates are then assembled and the gel cast as desired.

The only operative constraint that determines the size of a gel that is of use according to the invention is the physical ability of one of skill in the art to cast such a gel. The casting of gels of up to one meter in length is, while cumbersome, a procedure well known to workers skilled in nucleic acid sequencing technology. A larger gel, if produced, is also of use according to the invention. An extremely small gel is cut from a larger whole after polymerization is complete.

Note that at least one procedure for casting a polyacrylamide gel with bioactive substances, such as enzymes, entrapped within its matrix is known in the art (O'Driscoll, 1976, Methods Enzymol., 44: 169-183); a similar protocol, using photo-crosslinkable polyethylene glycol resins, that permit entrapment of living cells in a gel matrix has also been documented (Nojima and Yamada, 1987, Methods Enzymol., 136: 380-394). Such methods are of use according to the invention. As mentioned below, whole cells are typically cast into agarose for the purpose of delivering intact chromosomal DNA into a matrix suitable for pulsed-field gel electrophoresis or to serve as a "lawn" of host cells that will support bacteriophage growth prior to the lifting of plaques according to the method of Benton and Davis (see Maniatis et al., 1982, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). In short, electrophoresis-grade agarose (e.g., Ultrapure; Life Technologies/Gibco-BRL; is dissolved in a physiological (isotonic)

buffer and allowed to equilibrate to a temperature of 50° to 52°C in a tube, bottle or flask. Cells are then added to the agarose and mixed thoroughly, but rapidly (if in a bottle or tube, by capping and inversion, if in a flask, by swirling), before the mixture is decanted or pipetted into a gel tray. If low-melting point agarose is used, it may be brought to a much lower temperature (down to approximately room temperature, depending upon the concentration of the agarose) prior to the addition of cells. This is desirable for some cell types; however, if electrophoresis is to follow cell lysis prior to covalent attachment of the molecules of the resultant nucleic acid pool to the support, it is performed under refrigeration, such as in a 4° to 10°C 'cold' room.

Immobilization of nucleic acid molecules to the support matrix according to the invention is accomplished by any of several procedures. Direct immobilizing, as through use of 3'-terminal tags bearing chemical groups suitable for covalent linkage to the support, hybridization of single-stranded molecules of the pool of nucleic acid molecules to oligonucleotide primers already bound to the support or the spreading of the nucleic acid molecules on the support accompanied by the introduction of primers, added either before or after plating, that may be covalently linked to the support, may be performed. Where pre-immobilized primers are used, they are designed to capture a broad spectrum of sequence motifs (for example, all possible multimers of a given chain length, e.g., hexamers), nucleic acids with homology to a specific sequence or nucleic acids containing variations on a particular sequence motif. Alternatively, the primers encompass a synthetic molecular feature common to all members of the pool of nucleic acid molecules, such as a linker sequence (see above).

Oligonucleotide primers useful according to the invention are single-stranded DNA or RNA molecules that are hybridizable to a nucleic acid template to prime enzymatic synthesis of a second nucleic acid strand. The primer is complementary to a portion of a target molecule present in a pool of nucleic acid molecules used in the preparation of sets of arrays of the invention.

It is contemplated that such a molecule is prepared by synthetic methods, either chemical or enzymatic. Alternatively, such a molecule or a fragment thereof is naturally occurring, and is isolated from its natural source or purchased from a commercial supplier.

Oligonucleotide primers are 6 to 100, and even up to 1,000, nucleotides in length, but ideally from 10 to 30 nucleotides, although oligonucleotides of different length are of use.

Typically, selective hybridization occurs when two nucleic acid sequences are substantially complementary (at least about 65% complementary over a stretch of at least 14 to 25 nucleotides, preferably at least about 75%, more preferably at least about 90% complementary). See Kanehisa, M., 1984, Nucleic Acids Res. 12: 203, incorporated herein by reference. As a result, it is expected that a certain degree of mismatch at the priming site is tolerated. Such mismatch may be small, such as a mono-, di- or tri-nucleotide. Alternatively, it may encompass loops, which we define as regions in which mismatch encompasses an uninterrupted series of four or more nucleotides.

Overall, five factors influence the efficiency and selectivity of hybridization of the primer to a second nucleic acid molecule. These factors, which are (i) primer length, (ii) the nucleotide sequence and/or composition, (iii) hybridization temperature, (iv) buffer chemistry and (v) the potential for steric hindrance in the region to which the primer is required to hybridize, are important considerations when non-random priming sequences are designed.

There is a positive correlation between primer length and both the efficiency and accuracy with which a primer will anneal to a target sequence; longer sequences have a higher T_M than do shorter ones, and are less likely to be repeated within a given target sequence, thereby cutting down on promiscuous hybridization. Primer sequences with a high G-C content or that comprise palindromic sequences tend to self-hybridize, as do their intended target sites, since unimolecular, rather than bimolecular, hybridization kinetics are generally favored in solution; at the same time, it is important to design a primer containing sufficient numbers of G-C nucleotide pairings to bind the target sequence tightly, since each such pair is bound by three hydrogen bonds, rather than the two that are found when A and T bases pair. Hybridization temperature varies inversely with primer annealing efficiency, as does the concentration of organic solvents, e.g., formamide, that might be included in a hybridization mixture, while increases in salt concentration facilitate binding. Under stringent hybridization conditions, longer probes hybridize more efficiently than do shorter ones, which are sufficient under more permissive conditions. Stringent hybridization

conditions typically include salt concentrations of less than about 1M, more usually less than about 500 mM and preferably less than about 200 mM. Hybridization temperatures range from as low as 0°C to greater than 22°C, greater than about 30°C, and (most often) in excess of about 37°C. Longer fragments may require higher hybridization temperatures for specific hybridization. As several factors affect the stringency of hybridization, the combination of parameters is more important than the absolute measure of any one alone.

Primers are designed with the above first four considerations in mind. While estimates of the relative merits of numerous sequences are made mentally, computer programs have been designed to assist in the evaluation of these several parameters and the optimization of primer sequences. Examples of such programs are "PrimerSelect" of the DNAStar™ software package (DNAStar, Inc.; Madison, WI) and OLIGO 4.0 (National Biosciences, Inc.). Once designed, suitable oligonucleotides are prepared by a suitable method, e.g., the phosphoramidite method described by Beaucage and Carruthers (1981, Tetrahedron Lett., 22: 1859-1862) or the triester method according to Matteucci et al. (1981, J. Am. Chem. Soc., 103: 3185), both incorporated herein by reference, or by other chemical methods using either a commercial automated oligonucleotide synthesizer or VLSIPS™ technology.

Two means of crosslinking a nucleic acid molecule to a preferred support of the invention, a polyacrylamide gel sheet, will be discussed in some detail. The first (provided by Khrapko et al., 1996, U.S. Patent No. 5,552,270) involves the 3' capping of nucleic acid molecules with 3-methyl uridine; using this method, the nucleic acid molecules of the libraries of the present invention are prepared so as to include this modified base at their 3' ends. In the cited protocol, an 8% polyacrylamide gel (30:1, acrylamide: bis-acrylamide) sheet 30 μm in thickness is cast and then exposed to 50% hydrazine at room temperature for 1 hour; such a gel is also of use according to the present invention. The matrix is then air dried to the extent that it will absorb a solution containing nucleic acid molecules, as described below. Nucleic acid molecules containing 3-methyl uridine at their 3' ends are oxidized with 1 mM sodium periodate ($NaIO_4$) for 10 minutes to 1 hour at room temperature, precipitated with 8 to 10 volumes of 2% $LiClO_4$ in acetone and dissolved in water at a concentration of 10 pmol/μl. This concentration is adjusted so that when the nucleic acid

molecules are spread upon the support in a volume that covers its surface evenly, yet is efficiently (i.e. completely) absorbed by it, the density of nucleic acid molecules of the array falls within the range discussed above. The nucleic acid molecules are spread over the gel surface and the plates are placed in a humidified chamber for 4 hours. They are then dried for 0.5 hour at room temperature and washed in a buffer that is appropriate to their subsequent use. Alternatively, the gels are rinsed in water, re-dried and stored at -20°C until needed. It is said that the overall yield of nucleic acid that is bound to the gel is 80% and that of these molecules, 98% are specifically linked through their oxidized 3' groups.

A second crosslinking moiety that is of use in attaching nucleic acid molecules covalently to a polyacrylamide sheet is a 5' acrylyl group, which is attached to the primers used in Example 4. Oligonucleotide primers bearing such a modified base at their 5' ends may be used according to the invention. In particular, such oligonucleotides are cast directly into the gel, such that the acrylyl group becomes an integral, covalently-bonded part of the polymerizing matrix. The 3' end of the primer remains unbound, so that it is free to interact with- and hybridize to a nucleic acid molecule of the pool and prime its enzymatic second-strand synthesis.

Alternatively, hexaethylene glycol is used to covalently link nucleic acid molecules to nylon or other support matrices (Adams and Kron, 1994, U.S. Patent No. 5,641,658). In addition, nucleic acid molecules are crosslinked to nylon *via* irradiation with ultraviolet light. While the length of time for which a support is irradiated as well as the optimal distance from the ultraviolet source is calibrated with each instrument used, due to variations in wavelength and transmission strength, at least one irradiation device designed specifically for crosslinking of nucleic acid molecules to hybridization membranes is commercially available (Stratalinker; Stratagene). It should be noted that in the process of crosslinking *via* irradiation, limited nicking of nucleic acid strand occurs; however, the amount of nicking is generally negligible under conditions such as those used in hybridization procedures. Attachment of nucleic acid molecules to the support at positions that are neither 5'- nor 3'-terminal also occurs, but it should be noted that the potential for utility of an array so crosslinked is largely uncompromised, as such crosslinking does not inhibit hybridization of oligonucleotide primers to the immobilized molecule where it is bonded to the support. The

production of 'terminal' copies of an array of the invention, i.e. those that will not serve as templates for further replication, is not affected by the method of crosslinking; however, in situations in which sites of covalent linkage are, preferably, at the termini of molecules of the array, crosslinking methods other than ultraviolet irradiation are employed.

Step 3. Amplification of the Nucleic Acid Molecules of the Array

The molecules are amplified *in situ* (Tsongalis et al., 1994, Clinical Chemistry, 40: 381-384; see also review by Long and Komminoth, 1997, Methods Mol. Biol., 71: 141-161) by standard molecular techniques, such as thermal-cycled PCR (Mullis and Faloona, 1987, Methods Enzymol., 155: 335-350) or isothermal 3SR (Gingeras et al., 1990, Annales de Biologie Clinique, 48(7): 498-501; Guatelli et al., 1990, Proc. Natl. Acad. Sci. U.S.A., 87: 1874). Another method of nucleic acid amplification that is of use according to the invention is the DNA ligase amplification reaction (LAR), which has been described as permitting the exponential increase of specific short sequences through the activities of any one of several bacterial DNA ligases (Wu and Wallace, 1989, Genomics, 4: 560). The contents of this article are herein incorporated by reference.

The polymerase chain reaction (PCR), which uses multiple cycles of DNA replication catalyzed by a thermostable, DNA-dependent DNA polymerase to amplify the target sequence of interest, is well known in the art, and is presented in detail in the Examples below. The second amplification process, 3SR, is an outgrowth of the transcription-based amplification system (TAS), which capitalizes on the high promoter sequence specificity and reiterative properties of bacteriophage DNA-dependent RNA polymerases to decrease the number of amplification cycles necessary to achieve high amplification levels (Kwoh et al., 1989, Proc. Natl. Acad. Sci. U.S.A., 83: 1173-1177). The 3SR method comprises an isothermal, Self-Sustained Sequence Replication amplification reaction, is as follows:

Each priming oligonucleotide contains the T7 RNA polymerase binding sequence (TAATACGACTCACTATA [SEQ ID NO:1]) and the preferred transcriptional initiation site. The remaining sequence of each primer is complementary to the target sequence on the molecule to be amplified.

The 3SR amplification reaction is carried out in 100 µl and contains the target RNA, 40 mM Tris·HCl, ph 8.1, 20 mM MgCl2, 2 mM spermidine·HCl, 5mM dithiothreitol, 80 µg/ml BSA, 1 mM dATP, 1 mM dGTP, 1 mM dTTP, 4 mMATP, 4 mM CTP, 1 mM GTP, 4 mM dTTP, 4 mM ATP, 4 mM CTP, 4 mM GTP, 4 mMUTP, and a suitable amount of oligonucleotide primer (250 ng of a 57-mer; this amount is scaled up or down, proportionally, depending upon the length of the primer sequence). Three to 6 attomoles of the nucleic acid target for the 3SR reactions is used. As a control for background, a 3SR reaction without any target (H_2O) is run. The reaction mixture is heated to 100°C for 1 minute, and then rapidly chilled to 42°C. After 1 minute, 10 units (usually in a volume of approximately 2 µl) of reverse transcriptase, (e.g., avian myoblastosis virus reverse transcriptase, AMV-RT; Life Technologies/Gibco-BRL) is added. The reaction is incubated for 10 minutes, at 42°C and then heated to 100°C. for 1 minute. (If a 3SR reaction is performed using a single-stranded template, the reaction mixture is heated instead to 65°C for 1 minute.) Reactions are then cooled to 37°C for 2 minutes prior to the addition of 4.6 µl of a 3SR enzyme mix, which contains 1.6 µl of AMV-RT at 18.5 units/µl, 1.0 µl T7 RNA polymerase (both e.g., from Stratagene; La Jolla, CA) at 100 units/µl and 2.0 µl *E. Coli* RNase H at 4 units/µl (e.g., from Gibco/Life Technologies; Gaithersburg, MD). It is well within the knowledge of one of skill in the art to adjust enzyme volumes as needed to account for variations in the specific activities of enzymes drawn from different production lots or supplied by different manufacturers. The reaction is incubated at 37°C for 1 hour and stopped by freezing. While the handling of reagents varies depending on the physical size of the array (which planar surface, if large, requires containment such as a tray or thermal-resistant hybridization bag rather than a tube), this method is of use to amplify the molecules of an array according to the invention.

Other methods which are of use in the amplification of molecules of the array include, but are not limited to, nucleic acid sequence-based amplification (NASBA; Compton, 1991, Nature, 350: 91-92, incorporated herein by reference) and strand-displacement amplification (SDA; Walker et al., 1992, Nucleic Acids Res., 20: 1691-1696, incorporated herein by reference).

Additional Methods of Producing a Nucleic Acid Array

In addition to the methods presented above, alternative methods may be used in the production of nucleic acid arrays to generate arrays that might prove to be more advantageous for expressing protein arrays. These alternative methods include, but are not limited to, spray-painted arrays (inkjet), geometrical focusing, multiplex PCR, and amplification of nucleic acids in a polymer gel. These

Spray-Painted Arrays (Inkjet)

Immobilized nucleic acid molecules may, if desired, be produced using a device (e.g.,, any commercially-available inkjet printer, which may be used in substantially unmodified form) which sprays a focused burst of nucleic acid synthesis compounds onto a support (see Castellino, 1997, Genome Res., 7: 943-976). Such a method is currently in practice at Incyte Pharmaceuticals and Rosetta Biosystems, Inc., the latter of which employs what are said to be minimally-modified Epson inkjet cartridges (Epson America, Inc.; Torrance, CA). The method of inkjet deposition depends upon the piezoelectric effect, whereby a narrow tube containing a liquid of interest (in this case, oligonucleotide synthesis reagents) is encircled by an adapter. An electric charge sent across the adapter causes the adapter to expand at a different rate than the tube, and forces a small drop of liquid containing phosphoramidite chemistry reagents from the tube onto a coated slide or other support.

Reagents are deposited onto a discrete region of the support, such that each region forms a feature of the array; the desired nucleic acid sequence is synthesized drop-by-drop at each position, as is true in other methods known in the art. If the angle of dispersion of reagents is narrow, it is possible to create an array comprising many features. Alternatively, if the spraying device is more broadly focused, such that it disperses nucleic acid synthesis reagents in a wider angle, as much as an entire support is covered each time, and an array is produced in which each member has the same sequence (i.e. the array has only a single feature).

Arrays of both types are of use in the invention; a multi-feature array produced by the inkjet method is used in array templating, as described above; a random library of nucleic

acid molecules are spread upon such an array as a homogeneous solution comprising a mixed pool of nucleic acid molecules, by contacting the array with a tissue sample comprising nucleic acid molecules, or by contacting the array with another array, such as a chromosomal array (Example 7) or an RNA localization array (Example 6).

Alternatively, a single-feature array produced by the inkjet method is used by the same methods to immobilize nucleic acid molecules of a library which comprise a common sequence, whether a naturally-occurring sequence of interest (e.g., a regulatory motif) or an oligonucleotide primer sequence comprised by all or a subset of library members, as described herein above and in Example 4, below.

Nucleic acid molecules which thereby are immobilized upon an ordered inkjet array (whether such an array comprises one or a plurality of oligonucleotide features) are amplified *in situ*, transferred to a semi-solid support and immobilized thereon to form a first randomly-patterned, immobilized nucleic acid array, which is subsequently used as a template with which to produce a set of such arrays according to the invention, all as described above.

Geometrical Focusing

A characteristic of the replica amplification process is that each replica will tend to occupy a larger area than the feature from which it was made. This is because the feature molecules transferred to the replica may come from anywhere within the circumferential area occupied by the template feature. Subsequent amplification of the transferred molecules will necessarily increase the area occupied by the feature relative to that occupied by the template feature. It is clear that this phenomenon will limit the practical number of times an array may be sequentially replicated without contamination of surrounding features. There are several approaches to solving this problem.

First, as mentioned previously, more than one replica of an amplified array may be made per amplification. It is clear that the "earlier" in the replication process a given array is replicated, the less area its features will occupy relative to those made later. That is, the more replicas one can make of an original amplified array before re-amplifying the template, the more arrays with smaller features one will have. The number of replicas of a given array which may be made without re-amplification of the template may be determined empirically

by, for example, hybridization of a sequential series of amplified replicas from a single array with an oligonucleotide which hybridizes with a sequence common to every feature. Comparison of the hybridization signals from the first replica to those of subsequent replicas made from the same template without re-amplification of the template will indicate at what point features begin to be lost from the replicas.

Second, one may reduce the number of PCR cycles used in the amplification process. Because the amplification is exponential, a small change in the cycle number can have a profound influence on the area occupied by the feature. This will clearly not solve the problem completely, but when combined with the first approach it can extend the useful number of cycles of amplification and replication for a given array. The practical number of PCR cycles to use for each round of amplification may also be estimated empirically by making several replicas from a single template array without re-amplification, and then subjecting individual replicas in the series to increasing numbers of PCR cycles. For example, replicas may be subjected to 10, 20, and 30 amplification cycles, followed by hybridization with a fluorescent probe sequence common to all features of the array. Visualization of the hybridized array by fluorescence microscopy will indicate at which point the features begin to intrude upon one another. Clearly, the starting size of the feature will influence the number of PCR cycles allowable per replication cycle, but it is within the ability of one skilled in the art to determine generally how many cycles are optimal to obtain enough DNA for subsequent rounds of replica amplification without widespread contamination of surrounding features.

A third approach recognizes the fact that the amplified features occupy more than just the two dimensional area of the surface they sit upon. Rather, each amplified feature occupies a hemispherical space with a radius, r. If the features are situated on one slide, which for discussion will be designated the "bottom" slide, and covered by another slide (the "top" slide) set at a uniform, fixed distance from the bottom slide, one will note that as the hemispherical feature expands with rounds of amplification, the portion of the growing hemisphere which first contacts the top slide will be much smaller in cross-sectional area than the portion in contact with the bottom slide. This presents a smaller surface area, with all

sequence information intact, from which to make replicas that do not occupy greater surface area than their template features. This method will be referred to as "geometrical focusing."

For example, after 30 cycles in 15% polyacrylamide, 500 bp amplicons will form hemispheres with a 10 micron radius. The length of the template and the percentage of acrylamide in the gel influence the size of the amplified features such that, for a given number of cycles, the size of the features decreases as the length of the template or the percentage of acrylamide increases. In general, the size of an amplified feature with respect to a given number of amplification cycles under given conditions is determined empirically by visualizing it with a fluorescent confocal microscope or fluorimager after staining with a fluorescent intercalator. Labeled primers or nucleotides may also be used to "light up" the feature for measurement by this method.

The distance between the surface bearing the array and the surface the array is to be transferred to may be controlled using plastic spacers of the desired thickness along the edges of the slide. A small volume of polyacrylamide solution plus capillary action will take the volume out to the edges of a predetermined area of coverslip.

Another contemplated method of regulating or controlling the distance between surfaces in the geometrical focusing method involves the use of optical feedback, such as Newton rings or other interferometry, to adjust pressure locally across the surfaces. The adjustment may be accomplished by a scanning laser that heats a differential thermal expansion plate differentially based on the optical feedback.

As mentioned above, bioactive substances such as enzymes may be cast directly in polyacrylamide gels. Other reagents, including buffers and oligonucleotide primers may be either cast into the gels or added by diffusion or even electrophoretic pulses to the pre-formed gel matrices. If the upper plate has little or no adhesiveness to the gel (achieved, for example, through silane coating as described above), then when it is removed, the upper circle of each hemisphere is the only exposed DNA. Some of the exposed DNA can be transferred by microcontact printing using either plate, or by another round of polymerization from the upper plate. The radius of the circle exposed for transfer will be $c=\mathrm{sqrt}(r^2-d^2)$, where r is the radius of the hemisphere and d is the distance between the plates. Therefore, when r=10 microns and d=8 microns, the radius of the exposed circle, c=6 microns, less than

the size of the template feature. This exposed circle will thus have a cross-sectional area less than that occupied by the template feature, referred to as q, at the surface of the support. This slight reduction in the radius, and consequently the cross-sectional area of the transferred feature will work to keep the amplified replica features sharper through several rounds of replication. The distance between the plates may be 10%, 20%, 30%, 40%, on up to 50% or more less than the radius of the features being transferred. The surface area (of the support) occupied by the transferred features may be considered reduced or lessened if it is 10%, 20%, 30%, 40%, on up to approximately 80% less than the area occupied by features on the template array. The resolution of the features is considered to be preserved if the features remain essentially distinct after amplification of the transferred nucleic acid. It is noted that features which amplify with lower efficiency than others may be lost if the distance between plates is too large. Therefore, geometrical focusing will be most useful when combined with the other two approaches described for limiting the size of amplified replicas. That is, the number of replicas made from individual arrays early in the process should be maximized while the number of PCR cycles per amplification should be minimized.

Multiplex PCR

Multiplex PCR refers to the process of amplifying a number of different DNA molecules in the same PCR reaction. Generally, the process involves the addition of multiple primer pairs, each pair specific for the amplification of a single DNA target species. A major goal of investigators is to apply the power of multiplex PCR to the problem of high throughput genotyping of individuals for specific genetic markers. If 100,000 polymorphic markers are to be assayed per genome, it would be very expensive to perform 100,000 individual PCR reactions. Some advances have been made in multiplexing PCR reactions (Chamberlain et al., 1988, Nucl. Acids Res. 16:11141), and the degree of multiplexing of the PCR has been scaled up, followed by hybridization to an array of allele-specific probes (Wang et al., 1998, Science 280: 1077). However, in the studies by Wang et al., the percentage of PCR products that successfully amplified decreased as the number of PCR primers added to the reaction increased. When approximately 100 primer pairs were used, about 90% of the PCR products were successfully amplified. When the number of primer

pairs was increased to about 500, about 50% of the PCR products were successfully amplified.

The decreasing efficiency with increasing number of primers is due in large part to the phenomenon of "primer dimer" formation. Primer dimers are the result of fortuitous 3' terminal complementarity of 4 bp or more between primers. This complementarity allows hybridization which is stabilized by polymerase recognition and extension of both strands. After the first cycle of extension, the complementarity is no longer limited to the 3' terminal nucleotides; rather, the entire primer dimer is now complementary to the primers. This reaction efficiently competes with the desired amplification reaction, in part because the concentration of the primers is significantly greater than that of the desired amplification target, kinetically favoring the amplification of the primer dimers. This phenomenon increases with increasing numbers and concentrations of primers.

A new approach to solving these inherent problems with multiplex PCR uses microarrays of immobilized, amplified PCR primers. By immobilizing at least one of the PCR primers, the method reduces the possibilities for non-specific primer interactions. The local concentration of primers is high enough for amplification, yet the individual primers are restricted from interacting non-specifically with one another.

Another disadvantage of standard multiplex PCR is that individual primer pairs must be synthesized for each polymorphic target. Genotyping DNA with 100,000 polymorphism targets would require, in theory, 200,000 different PCR primers. Not only is the synthesis of such primers costly and time consuming, but not all primer designs succeed in producing a desired PCR product. Therefore considerable time and energy will be spent optimizing the primer designs.

According to the new multiplex PCR method, one of the primers has a 5' end which is generic for the entire multiplex PCR reaction, such that the entire multiplex reaction will have that segment on the "mobile" primer. This 5' generic sequence may contain a restriction site for later cloning, a bacteriophage or other promoter for transcription of the products, or some other useful or identifiable sequence. The 3' end of the mobile primer is complementary to any genomic (or cDNA) sequence which is to be amplified at a reasonable PCR distance from the 3' end of the immobile primer. In other words, the 3' end of the mobile primer is

randomized. The length of the randomized 3' sequence may be as few as 5 nucleotides, up to 10 nucleotides or more. The second, or "specific" primers are immobilized (according to methods known in the art or described herein) to keep them from diffusing into the other primer pair zones while the mobile primer allows the extended product to diffuse.

There are at least two ways primer pairs may be distributed. First, two presynthesized Acrydite primers may be codeposited (Kenney et al., 1998, Biotechniques 25: 516-521; Rehman et al., 1999, Nucl. Acids Res. 27: 649-655), along with template and polymerase, in a gel volume element, for example by aerosol, emulsion, or inkjet printer, from an equimolar primer mixture. Alternatively, the primers may be derived from genomic DNA by a localized PCR. Generic primers can be used with one immobilized primer to make amplified features, and then release the new extended primers by exonuclease or type II restriction enzymes as described elsewhere herein. The new extended primers would then be copolymerized, along with template and polymerase, into the gel.

The process of this modified multiplex PCR method can be thought of as essentially two different steps. In the first, primers immobilized in a microarray hybridize with their complementary sequence in the template and are extended. In the second, and subsequent steps, the 3' (randomized) end of the mobile primers hybridizes at some point along the length of the extended immobilized primer and is itself extended. In subsequent cycles, other molecules in the immobilized primer features hybridize with the products of the previous extension, allowing extension, and so on, yielding exponential amplification as in standard PCR.

The multiplex PCR strategy need not involve replica printing.

Amplification of Nucleic Acid Molecules in a Polymer Gel

According to one aspect of the present invention, an array of nucleic acid molecules is produced as a result of amplification of an initial nucleic acid molecule, whether alone or as part of a plasmid, in a polymer gel or other suitable gel matrix which is placed on a solid support. The gel matrix advantageously serves to immobilize the amplified nucleic acid molecules whether by covalent interaction or steric hindrance between the nucleic acid molecules and the gel matrix. Suitable gel matrices within the scope of the present invention

include those prepared by polymerization of one or more commercially available monomers such as acrylamide and the like to form a polyacrylamide gel matrix. One of ordinary skill in the art will readily recognize that other suitable polymer-based matrices are useful in the practice of the present invention. The present invention also includes other gel matrices such as those made from starches, agarose and the like. As an illustration of one aspect of the present invention, polyacrylamide gel matrices will be discussed.

The solid support can be fashioned of any material known to those of skill in the art to be suitable in the practice of the present invention. The surface of the solid support can optionally be pretreated in a manner to increase adherence of the polyacrylamide gel to the solid support. According to a preferred embodiment, the solid support is fashioned out of glass. A convenient solid support for use with the present invention is a glass microscope slide.

According to a general embodiment of the present invention, acrylamide monomers are polymerized in a liquid mixture containing at least one standard commercially available or readily manufactured oligonucleotide primer reagent, such as a PCR primer, and an effective amount of template nucleic acid. One of ordinary skill in the art will recognize that the principles of the present invention apply to single stranded nucleic acids, double stranded nucleic acids, or triple stranded nucleic acids. For purposes of illustration of the present invention, template DNA and PCR reagents will be discussed. According to one embodiment, the PCR primers are present in pairs (at least two) and in amounts sufficient to amplify the DNA template when subject to certain reaction conditions. The resulting gel matrix is poured onto a solid support which is subjected to conditions sufficient to effect amplification of the DNA template. As the amplification reaction proceeds, the products remain localized near their respective templates due in part to the polyacrylamide gel. The amplification reaction results in an amplified sequence feature consisting of 10^8 or more essentially identical molecules.

According to one aspect of the present invention, one or more of the PCR primers includes a linker moiety which covalently reacts with the chosen monomer during polymerization of the gel matrix. As a result, the PCR primers become covalently bound to and immobilized within the polymer gel matrix. One such linker moiety for use with

polyacryamide gel matrices includes a commercially available linker moiety known as ACRYDITE. ACRYDITE is a phosphoroamidite that contains an ethylene group which enters into a free-radical copolymerization with acrylamide. A PCR primer can be modified to include the ACRYDITE moiety at the 5' end (Kenney et al., 1998, BioTechniques 25: 516-521). As a result, the amplified DNA in each feature can be covalently attached by one of its ends to the polyacrylamide gel matrix. One of ordinary skill in the art will become aware of other linker moieties useful in the present invention to covalently bind to the gel matrix of choice based upon the disclosure presented herein.

Primers

Primers useful in the practice of the present invention were obtained from Operon (CA) and are identified below. Certain primers used for creation of cassettes had common sequences which are indicated below by bold type, italicized type, underscored type, or bold-italicized type.

Primers used for solid phase amplification:

Primer OutF 5'-**cca cta cgc ctc cgc ttt cct ctc** -3' (SEQ ID NO:2)

Primer OutR 5'-*ctg ccc cgg gtt cct cat tct ct*-3' (SEQ ID NO:3)

Primer AcrOutF 5'-Q**cca cta cgc ctc cgc ttt cct ctc**-3' (SEQ ID NO:4)

Primer InF 5'-ggg cgg aag ctt gaa gga ggt att-3' (SEQ ID NO:5)

Primer InR 5'-***gcc cgg tct cga gcg tct gtt ta***-3' (SEQ ID NO:6)

Primer AcrInF 5'-Qggg cgg aag ctt gaa gga ggt att-3' (SEQ ID NO:7)

Primer PucF:

5' - ggg cgg aag ctt gaa gga ggt att taa gga gaa aat acc gca tca gg-3' (SEQ ID NO:8)

Primer PucR1:

5'- ***gcc cgg tct cga gcg tct gtt t***ac acc gat cgc cct tcc caa ca-3' (SEQ ID NO:9)

Primer PucR2:

5'- ***gcc cgg tct cga gcg tct gtt t***aa att cac tgg ccg tcg ttt tac aa-3' (SEQ ID NO:10)

Primer PucR3:

5'- ***gcc cgg tct cga gcg tct gtt t***ac caa tac gca aac cgc ctc tcc - 3' (SEQ ID NO:11)

Primer PucNestF:

5'- **cca cta cgc ctc cgc ttt cct ctc** ggg cgg aag ctt gaa gga ggt att-3'(SEQ ID NO:12)

Primer PucNestR:

5'-*ctg ccc cgg gtt cct cat tct* ***ctg ccc ggt ctc gag cgt ctg ttt a***-3'(SEQ ID NO:13)

The primers AcrOutF and AcrInF include an ACRYDITE modification which is commercially available from Mosaic Technologies, Inc. (Waltham, MA, USA). The primers are modified at their 5' ends with the ACRYDITE moiety which is designated by the character Q in the sequences listed above. Since ACRYDITE is a phosphoramidite that contains an ethylene group capable of free-radical copolymerization with acrylamide, primers including the ACRYDITE moiety will polymerize directly into and become covalently bound to the acrylamide gel as it solidifies (Kenney et al., 1998, supra).

Design of Amplification Cassettes

Amplification cassettes useful in the practice of the present invention were prepared. The plasmid pUC19 was amplified in a PCR reaction according to the following method. 50 µl of a PCR mixture containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 0.01% gelatin, 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.5 µM primer PucF, 0.5 µM primer PucR2, 2 ng pUC19 plasmid, and 2 units Taq (Sigma) was cycled in an MJ Research PTC-100 thermocycler. The cycle used was denaturation (1 min at 94°C), 5 cycles (10 sec at 94°C, 10 sec at 55°C, 1min at 72°C), 20 cycles (10 sec at 94°C, 1 min at 68°C), and extension (3 min at 72°C). The PCR product was purified using Qiaquick PCR purification columns (Qiagen), and resuspended in deionized water.

Two additional amplification cassettes were created, a 120 bp cassette (CP-120) and a 514 bp cassette (CP-514), and used to determine the relationship between the length of the amplification cassette and the resulting amplified feature diameter. These two cassettes were created as described above, except the reverse primers PucR1 and PucR3 were used instead of PucR2 in the first PCR mixture.

A further additional 281 bp cassette (CP-281) was also created and used in replica amplification experiments. CP-281 is identical to CP-234 expect that it is flanked by two additional primer sites. These primer sites allowed a nested solid phase PCR reaction to

create duplicate amplified feature slides without contamination from primer-dimer molecules. CP-218 was created by cycling a PCR mixture of 10 ng CP-234, 10 mM Tris-HCl pH 8.3, 50 mM KCl, 0.01% gelatin, 1.5 mM $MgCl_2$, 200 μM dNTP's, 0.5 μM primer PucNestF, 0.5 μM primer PucNestR, and 2 units Taq (Sigma) as follows: denaturation (1 min at 94°C), 5 cycles (10 sec at 94°C, 10 sec at 55°C, 1min at 72°C), 22 cycles (10 sec at 94°C, 1 min at 68°C), and extension (3 min at 72°C). The PCR product was purified using Qiaquick PCR purification columns (Qiagen), and resuspended in deionized water.

Creating Slides of Nucleic Acid Molecules Immobilized in a Gel Matrix

One aspect of the present invention includes a method of making an array of nucleic acid molecules that are immobilized in a gel matrix. According to the present invention, a liquid mixture of template DNA, a pair of PCR primers, at least one of which primers is optionally 5' ACRYDITE modified, and acrylamide monomers is prepared. The liquid mixture is poured onto a solid substrate such as a glass slide. The liquid mixture is then polymerized under suitable conditions. The template DNA is also amplified by PCR under suitable conditions. The result is an array having amplified nucleic acid molecules that are immobilized. The method is described in greater detail in the following non-limiting example.

To create an array slide according to this aspect of the invention, template DNA was amplified by PCR in a polyacrylamide gel poured onto a glass microscope slide. Dilute amounts of template CP-234 (0-360 molecules, quantified by ethidium bromide staining and gel electrophoresis) were added to the solid phase PCR mixture containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.01% gelatin, 1.5 mM $MgCl_2$, 200 μM dNTP's, 0.5 μM primers, 2 ng pUC19 plasmid, 10 units JumpStart Taq (Sigma), 6% Acrylamide, 0.32% Bis-Acrylamide, 1 μM primer AcrInF, and 1 μM primer InR. Two 65 μl frame-seal chambers (MJ research) were attached to a glass microscope slide that had been pre-treated with bind-silane (Pharmacia). Other types of bind-silane are commercially available from Sigma. Pre-treatment of a glass slide with bind-silane results in the enhanced binding of the polymerized polyacrylamide to the slide.

2.5 μl of 5% ammonium persulfate, and 2.5 μl of 5% TEMED were added to 150 μl of the solid phase PCR mixture. 65 μl of this solution was added to each chamber. The chambers were then immediately covered with No. 2 coverslips (Fisher, 18 mm x 18 mm), and the gel matrix was allowed to polymerize for 10-15 minutes. Thermostable, template-dependent DNA polymerases other than JumpStart Taq polymerase are known to those skilled in the art and are also useful in this, and other aspects of the invention.

The slide was then cycled using a PTC-200 thermal cycler (MJ Research) adapted for glass slides (16/16 twin tower block). The following program was used: denaturation (2 min at 94°C), 40 cycles (30 sec at 93°C, 45 sec at 62°C, 45 sec at 72°C), extension (2 min at 72°C). The coverslips were removed and the gels were stained in SYBR green I (diluted 5000 fold in TE, pH 8.0), and imaged on a Storm phosphorimager (Molecular Dynamics) or a confocal microscope (Leica).

Determining Relationship Between Amplified
<u>Feature Diameter, Template Length, and Acrylamide Concentration</u>

The relationship between amplified feature diameter, template length and acrylamide concentration was determined as follows. Slides were poured in the manner described above. The ratio of bis-acrylamide to acrylamide was 1:19 for all slides poured. After the slides were cycled, the coverslips were removed and the gels were stained as above. The gels were imaged using the Storm phosphorimager. Any gels with amplified features less than 300 μm in diameter were imaged on the confocal microscope. Care was taken to image only the amplified features that could be completely resolved from other amplified features. These images were captured, and the intensity values saved as a text file. The data were smoothed using a 17 point averaging algorithm, and the full width at half maximum of each amplified feature was recorded as its diameter.

Features of a DNA array were amplified on a glass microscope slide by performing solid phase PCR (see Lockley et al., 1997, Nucl. Acids Res. 25: 1313-1314) in an acrylamide gel. The general design of the template DNA cassettes used to create the amplified feature array slide is shown in Figure 7. The template DNA includes binding sites for the pair of PCR primers, one on either side of a sequence of interest. For most applications, the

sequence of interest will be a variable region, with the variable region of each cassette molecule containing a different DNA fragment. This complex template library will contain sequences derived from the genome or cDNA of the organism of interest flanked by constant regions that allow PCR amplification (Singer et al., 1997, Nucl. Acids Res. 25: 781-786). However, to demonstrate and optimize the *in vitro* cloning of DNA, only one species of DNA was used in the solid phase PCR: the cassette CP-234, a 234 base pair template derived from the plasmid pUC19. Very dilute amounts of the template DNA CP-234 were included in a PCR mix that contained 6% acrylamide and 0.3% bis-acrylamide. This mix was then used to pour a thin (250 μm) acrylamide gel on top of a glass microscope slide. One of the primers included in the mix contained an ACRYDITE group at its 5' end, so that it was immobilized in the acrylamide matrix when the gel polymerized. Solid phase PCR (so named because one of the primers is immobilized to a solid support) was performed by thermal cycling of the slide. The gels went through 40 cycles of denaturation, annealing and extension, and were stained using SYBR Green I.

Upon imaging, green fluorescent spheres were seen in the gels that had been poured with template DNA (Figure 8A). These spheres were not seen in the control slide lacking template DNA. The spheres were uniform in shape and roughly 300 μm in diameter, with little variation in size. The number of fluorescent spheres shows a linear dependence on the number of template molecules added (Figure 8B).

In order to confirm that the fluorescent spheres were DNA features which were amplified from a single molecule of the template cassette CP-234, stained spheres were removed using a toothpick and placed into a tube containing a PCR mixture, and the mix was thermal cycled. As a negative control, regions of the gel that did not contain fluorescent spheres were also removed using a toothpick, mixed with a PCR mixture and thermal cycled. The reactions were then run out on an agarose gel. The results are shown in Figure 8C. The sample containing the stained spheres clearly showed products at 234 bp as expected, while the sample containing regions of the gel that showed no spheres yielded no product.

While not wishing to be bound by any scientific theory, it is believed that the stained spheres shown in Figure 8A are due to the amplification of single template molecules. First, the number of amplified features obtained in each reaction is linearly dependent on the

amount of template included. As seen in Figure 8B, eighty percent of the template molecules added to each reaction yielded amplified features. Less than one hundred percent efficiency is believed to be due to possible damage to template molecules by the free radicals generated during the acrylamide polymerization, loss of template molecules to abstraction by tube or pipette tip walls, or the amount of template may have been underestimated when quantified by ethidium bromide staining. Second, amplified feature-picking experiments confirmed that product of expected length can be produced. Third, as shown in Figure 4, amplified feature size is strongly dependent on the length of the template.

In some experiments, a few larger fluorescent spheres (1-2 mm in diameter) were observed. Because these spheres were also observed on slides that were poured without template DNA, it was suspected that these spheres were the result of primer-primer mispriming (primer dimer). This was confirmed by repeating the sphere-picking experiment described above on the putative primer-dimer spheres (data not shown). Primer dimer spheres or features can be reduced or eliminated by raising the annealing temperature of the PCR and/or by careful primer design as known by those skilled in the art.

Because the number of amplified features per slide goes up with the inverse square of the feature size, it is necessary to minimize the size of each amplified feature in order to obtain slides with as many amplified features as possible. In order to determine the parameters that influence amplified feature size, solid phase PCR reactions were performed using template cassettes of different lengths. Acrylamide concentration was also varied. The results are shown in Figure 9.

The results, shown in Figure 9A, show that amplified feature radius decreases as template length increases and as the acrylamide percentage increases. Using the 514 base pair template, CP-514, and an acrylamide concentration of 15%, the amplified features produced were very small (average radius of 12.5 μm), and of uniform size (standard deviation of 0.29 μm).

These results showed that amplified feature radius was very sensitive to length of the template. In order to further minimize amplified feature size, a template cassette was created that was 1009 base pairs long. When this cassette was used as template in a solid phase PCR in 15% acrylamide, the resulting amplified features had radii of approximately 6 μm (Figure

9B). At this size, it is estimated that 5 million distinguishable amplified features can be poured on a single slide based on over 13.5 million being actually poured on the slide but that 63% of these will overlap one another. It is believed that amplified feature radius could be further reduced by increasing the length of the template DNA, by using fewer cycles of PCR, or by immobilizing both primers.

A simulation of amplified feature growth was developed to investigate the apparent relationship shown in Figure 4A between feature size and variation in size. This model assumes that at each cycle in the PCR reaction, every DNA molecule will move in a stochastic fashion (due to thermal energy) and then give rise to a complementary strand. The probability that a given molecule will give rise to a complementary strand is dependent on the number of unextended primers and the number of complementary strands in the immediate vicinity of the DNA. This model was tested using a number of different probability distribution functions for DNA motion with all runs being assumed that the DNA does not travel too far in relation to the average distance between immobilized primers. In all cases the results were qualitatively similar. This model predicts that template amplification in each feature is exponential during the early amplification cycles. As the amplified feature grows, it will reach a certain radius, the critical radius, after which the amplification proceeds at a polynomial rate. The critical radius is dependent on the diffusion coefficient of the template molecule, and the probability that a given DNA molecule is replicated after one cycle of the solid phase PCR. While not wishing to be bound by any one theory, one possible explanation is that one of the primers in the reaction is immobilized. Therefore, for an amplified feature to achieve exponential amplification, one strand of each full length DNA product in the feature must diffuse and anneal to an immobilized primer at each round of amplification. In this theory, during the early rounds, most of the immobilized primers in the vicinity of a template have not yet been extended, so the total number of DNA molecules in a feature increases exponentially with the cycle number. However, at later rounds, the DNA at the center of the feature cannot diffuse far enough to find immobilized primer that has not yet been extended. So, only the DNA near the circumference of the feature can continue to amplify. Therefore, the number of new DNA molecules generated

with each cycle increases as the square of the cycle number, so that the total number of DNA molecules in the feature increases with the cube of the cycle number.

Accordingly, it is possible, for example, that when the long DNA template, CP-514, was amplified to form amplified features, the features reached their critical radii and then grew very slowly for the rest of the reaction. Therefore, all of the amplified features tended to be the same size. In contrast, it is also possible that when the short DNA template, CP-120, was used, the features never reached their critical radii, so that some amplified features were bigger or smaller than others due to the stochastic nature of PCR.

B. Expression of a Nucleic Acid Array to Produce a Protein Array of the Invention

Once a nucleic acid array has been produced, peptides can be generated from the nucleic acid arrays. Applying the method described in Roberts, R.W. and Szostak, J.W. Proc. Natl. Acad. Sci. (1997), 94(23):12297-12302 "RNA-Peptide Fusions for the In Vitro Selection of Peptides and Proteins," hereby incorporated by reference in its entirety for all purposes, an array of nucleic acids that contain mRNAs is used to produce proteins which are attached by a stable covalent linkage to the mRNAs that respectively encode the proteins. In this manner, an array of immobilized proteins that are covalently attached to their encoding mRNAs is produced.

In general, according to the method of Roberts and Szostak, an mRNA is fused to a linker nucleic acid that contains the translational inhibitor puromycin attached at its 3' end. As exemplified by Roberts and Szostak, the nucleic acid may be a RNA-DNA hybrid molecule, such that 5' portion of the molecule is RNA with a 3' oligodeoxynucleotide linker that has a 3'-terminal puromycin. The linker serves to pause a translating ribosome, enabling the 3' puromycin to enter the ribosome and be attached to the nascent peptide chain. A pool of nucleic acid templates is translated in vitro, to produce nucleic acid-peptide fusions which are covalently linked to the mRNA. In the present invention, the mRNAs are an array of nucleic acids which are translated in vitro to produce an array of proteins linked to their encoding mRNAs.

Following the methods described by Roberts and Szostak, puromycin is first coupled to a controlled pore glass (CPG) solid support for synthesis of 3' puromycin nucleic acid templates. Puromycin$(HCl)_2$ is converted to the free base by dissolution in water, mixing with basic carbonate buffer, and extraction into chloroform. *N*-trifluoroacetyl puromycin is made by mixing the dried free base in 50/50 (vol/vol) dry pyridine (Fluka)/acetonitrile (Millipore) with an excess of trifluoroacetic anhydride (Fluka) for 1 hour at 25°C followed by workup with dilute ammonium hydroxide. *N*-trifluoroacetyl 5'-dimethoxytrityl (DMT) puromycin is made by using DMT-Cl (Sigma) (Jones, R. A., 1984, in *Oligonucleotide Synthesis: A Practical Approach*, ed. Gait, M. J. (IRL, Oxford), pages 23-34), and is attached to an aminohexyl CPG (Sigma) support through the 2' OH using a standard protocol for attachment of DNA through its 3' OH (Atkinson, T. and Smith, M., 1984, in *Oligonucleotide Synthesis: A Practical Approach*, ed. Gait, M. J. (IRL, Oxford), pages 35-81) with the exception that the coupling step is carried out in the presence of approximately 50 μmol activated puromycin per gram of CPG.

CPG-puromycin is then used as a solid support for automated synthesis of nucleic acids according to standard protocols for DNA and RNA synthesis (Millipore). Following synthesis, nucleic acids are deprotected in concentrated NH_4OH plus 25% vol/vol ethanol for 12 hours at 55°C and dried. The nucleic acids are gel purified on denaturing urea polyacrylamide gel electrophoresis, electroeluted using an Elutrap (Schleicher and Schuell), and desalted on NAP-25 columns (Pharmacia). The nucleic acids are arrayed on solid supports as described above in Example 1A. Proteins are produced from the nucleic acids by in vitro translation with reticulocyte lysate (Novagen) according to the manufacturer's specifications.

Another means for expressing proteins from arrayed mRNAs such that the proteins are attached to their respective encoding mRNAs is ribosome display, as described in Hanes et al., Nat. Biotechnol. (2000) 18:1287-92 "Picomolar Affinity Antibodies from a Fully Synthetic Naive Library Selected and Evolved by Ribosome Display," and Mattheakis et al., Proc. Natl. Acad. Sci. (1994), 91:9022-9026 "An In Vitro Polysome Display System for Identifying Ligands from Very Large Peptide Libraries," each of which are hereby incorporated by reference in their entirety for all purposes. In ribosome display technology, a

ribosomal complex is formed between an mRNA, a ribosome, and an encoded protein, such that the ribosome is stalled on the mRNA at the end of translation and the encoded protein is tethered to the ribosome, thereby non-covalently linking the mRNA to the encoded protein through the ribosome. The encoded protein is preferably tethered to the ribosome by a 20-30 amino acid spacer at the C terminus of the protein to enable the encoded protein to completely emerge from the ribosome and fold into its native conformation without steric hindrance from the ribosome. The mRNA construct is devoid of stop codons to maintain the covalent bond of the last amino acid of the peptide to the transfer RNA (tRNA), causing the ribosome to remain complexed with the mRNA.

Accordingly, ribosome display is used to create an array of proteins that are expressed from and non-covalently attached to an array of nucleic acids as described by Hanes et al. and Mattheakis et al.. The nucleic acids preferably are mRNAs containing a 3' spacer encoding a C terminus spacer that is attached to the functional protein encoded by the mRNA. The nucleic acids can be synthesized by any of a number of techniques known in art, including standard protocols for automated synthesis of RNA (Millipore) or enzymatic synthesis of RNA from a cloned DNA template expressed under the control of an appropriate promoter, such as a T7 promoter that is used with a T7 RNA polymerase transcription system (Promega). The nucleic acids are arrayed as described above in Example 1A.

The array of nucleic acids are translated in vitro using an *E. coli* S-30 system as described by Chen and Zubay (Chen, H.Z. and Zubay, G., 1983, *Methods. Enzymol.* 101, 674-690) and modified by Hanes et al. Translation is carried out for 10 minutes at 37°C in a reaction mixture containing 50 mM Tris-HOAc, ph 7.5, 30 mM NH_4Oac, 12.3 mM $Mg(OAC)_2$, 0.35 mM each amino acid, 2 mM ATP, 0.5 mM GTP, 1 mM cAMP, 0.5 mg/ml *E. coli* tRNA, 20 μg/ml folinic acid, 100 mM KOAc, 20 mM acetylphosphate, 1.5% polyethylene glycol 8000, 33 μg/ml rifampicin, 1 mg/ml vanadyl ribonucleoside complexes (VRC), and *E. coli* MRE600 extract (Chen and Zubay). Translation is stopped by adding $Mg(OAc)_2$ to a final concentration of 50 mM and cooling on ice.

Following translation of nucleic acid arrays as described above, the expressed, immobilized proteins can be modified either covalently or non-covalently according to standard methods well known in the art. Modifications of expressed proteins include, but are

not limited to, those derived by phosphorylation, glycosylation, proteases, chaperones, detergents, heat or solvent denaturation, metals, ions, as well as organic, inorganic, and organometallic compounds.

EXAMPLE 2

Replication of Arrays of the Invention

Following the production of a first nucleic acid array, as described above in Example 1, multiple copies of an array can optionally be made through replication of the array. Multiple copies of a nucleic acid array offer several uses and advantages over single copy arrays. For instance, multiple copies of an array enable replica-destructive methodologies, such as that described below, to be used with arrays. Also, multiple copy arrays can be used to reproducibly express proteins or alternatively, to readily bind protein. In this manner, multiple copy nucleic acid arrays can advantageously be used rather than protein arrays for storage of protein information, given that DNA is inherently more stable and less susceptible to degradation than protein. Multiple copies of a particular nucleic acid array thus can be expressed to enable performance of various protein-based assays both in parallel and reproducibly.

Replication of an Array

a. A master plate generated according to steps 1 through 3 of Example 1 is replica-plated by any of a number of methods (reviewed by Lederberg, 1989, Genetics, 121(3): 395-9) onto similar gel-chips. This replica is performed by directly contacting the compressible surfaces of the two gels face to face with sufficient pressure that a few molecules of each clone are transferred from the master to the replica. Such contact is brief, on the order of 1 second to 2 minutes. This is done for additional replicas from the same master, limited only by the number of molecules post-amplification available for transfer divided by the minimum number of molecules that must be transferred to achieve an acceptably faithful copy. While it is theoretically possible to transfer as little as a single molecule *per* feature, a more conservative approach is taken. The number of each species of molecule available for transfer never approaches a value so low as to raise concern about the

probability of feature loss or to the point at which a base substitution during replication of one member of a feature could, in subsequent rounds of amplification, create a significant (detectable) population of mutated molecules that might be mistaken for the unaltered sequence, unless errors of those types are within the limits of tolerance for the application for which the array is intended. Note that differential replicative efficiencies of the molecules of the array are not as great a concern as they would be in the case of amplification of a conventional library, such as a phage library, in solution or on a non-covalently-bound array. Because of the physical limitations on diffusion of molecules of any feature, one which is efficiently amplified cannot 'overgrow' one which is copied less efficiently, although the density of complete molecules of the latter on the array may be low. It is estimated that 10 to 100 molecules *per* feature are sufficient to achieve fidelity during the printing process. Typically, at least 100 to 1000 molecules are transferred.

Alternatively, the plated DNA is reproduced inexpensively by microcontact printing, or μCP, (Jackman et al, 1995, Science, 269(5224): 664-666, 1995) onto a surface with an initially uniform (or patterned) coating of two oligonucleotides (one or both immobilized by their 5' ends) suitable for *in situ* amplification. Pattern elements are transferred from an elastomeric support (comparable in its physical properties to support materials that are useful according to the invention) to a rigid, curved object that is rolled over it; if desired, a further, secondary transfer of the pattern elements from the rigid cylinder or other object onto a support is performed. The surface of one or both is compliant to achieve uniform contact. For example, 30 micron thin polyacrylamide films are used for immobilizing oligomers covalently as well as for *in situ* hybridizations (Khrapko, et al., 1991, DNA Sequence, 1(6):375-88). Effective contact printing is achieved with the transfer of very few molecules of double- or single-stranded DNA from each sub-feature to the corresponding point on the recipient support.

b. The replicas are then amplified as in step 3 of Example 1.

c. Alternatively, a replica serves as a master for subsequent array replications, limited by the diffusion of the features and the desired feature resolution.

Duplicating Array Slides

One aspect of the invention encompasses a method of making a plurality of arrays from a single array having nucleic acid molecules immobilized in a polyacrylamide gel. According to the method of the present invention, a liquid mixture of template DNA, a pair of PCR primers, at least one of which primers is 5' ACRYDITE modified, and acrylamide monomers is poured onto a solid substrate, such as a glass microscope slide, and then polymerized under suitable conditions to form a first layer. A liquid mixture of a pair of PCR primers, at least one of which primers is optionally 5' ACRYDITE modified, and acrylamide monomers without template DNA is poured on top of the first layer, and then polymerized to form a second layer. The template DNA is then amplified under suitable conditions to generate a nucleic acid array which is immobilized in the polyacrylamide gel matrix. Because the second layer is held in contact with the first layer during the amplification, a portion of the amplified nucleic acids from the first layer are transferred to the second layer whether by diffusion, adhesion, covalent bonding or other mechanism. The second layer is then removed and the process repeated as many times as desired to generate a plurality of arrays. The method is described in greater detail in the following non-limiting example.

To duplicate arrays of the present invention containing immobilized nucleic acids, a sandwich of two layers of acrylamide, the "transfer layer" and the "readout layer" is prepared. To create the transfer layer, template DNA is added to a solid phase PCR mix (10 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.01% gelatin, 1.5 mM $MgCl_2$, 200 μM dNTP's, 0.5 μM primers, 2 ng pUC19 plasmid, 10 units JumpStart Taq (Sigma), 6% Acrylamide, 0.32% Bis-Acrylamide, 1 μM primer AcrOutF, 1 μM primer OutR). Ten microliters of this solution are then pipetted onto a clean coverslip (18mm x 18mm), and the coverslip is picked up by a bind-silane treated slide. The slide is placed in an argon atmosphere to promote polymerization of the acrylamide. The coverslip is then removed, leaving a gel that is approximately 32 μm thick. To pour the readout layer, a fresh solid phase PCR mix is made; however, no template is added to this mixture. A frame seal chamber is then placed over the transfer layer, and, using a bind-silane treated glass coverslip, the readout layer (250 μm) is poured over the 32 μm transfer layer. The slide is then thermal cycled as described above.

When the coverslip is carefully removed from the top of the frame seal chamber, the readout layer will stick to the coverslip, while the transfer layer will be left on the slide. The readout layer can then be stained with SYBR Green I and imaged. The transfer layer is then used to make duplicates. To do so, the slide is washed 2x in 10 mM Tris-HCl, 2x in 500 mM KCl, 2x in 10 mM Tris, 100 mM KCl, and 2x in dH20. The duplicate gel is then made by placing a frame seal chamber (15 mm x 15mm) over the transfer layer, and pipetting 65 µl of the duplicate solid-phase PCR mix (10 mM Tris-HCl pH 8.3, 50 mM KCl, 0.01% gelatin, 1.5 mM $MgCl_2$, 200 µM dNTP's, 0.5 µM primer AcrInF, 0.5 µM primer InR, 10 units JumpStart Taq (Sigma), 6% Acrylamide, 0.32% Bis-Acrylamide), onto the transfer layer. The duplicate slide is then cycled as follows: denaturation (2 min at 94°C), 25 cycles (30 sec at 93°C, 45 sec at 62°C, 45 sec at 72°C), extension (2 min at 72°C). Because the coverslip used to pour the duplicate gel was not treated with bind-silane, the gel stuck to the transfer layer when the coverslip was removed; therefore when the duplicate was stained and imaged, the amplified feature pattern of the array was rotated 180 degrees from that of the readout layer.

According to the above protocol, a DNA array slide was created by pouring a thin, 3.1 µm gel containing template DNA (the template or transfer layer) on a bind silane-treated glass microscope slide, and then pouring a thicker gel (250 µm) over it, the thicker gel lacking template DNA but containing primers. When the sandwich is thermal cycled, the DNA in the thin layer produces amplified DNA features that span the interface between the two gels.

When the coverslip was carefully removed from the microscope slide, the thick gel remained intact and attached to the coverslip. This gel was stained with SYBR Green I and saved for comparison with the duplicate. Because the surface of the slide was treated with bind silane before the original was poured, the 3.1 µm layer of acrylamide (the template layer) remained bound to the surface of the slide. The slide was washed, and a new gel, the "duplicate," was poured on this glass slide. The duplicate was then thermal cycled and stained.

Figure 10 shows the imaged original slide (A) and duplicate amplified feature slide (B). The duplicate slide exhibited an amplified DNA feature pattern that is identical to that of

the original. The amplified DNA features on the duplicate tend to be slightly larger than those on the original due to diffusion in the duplicate solid phase PCR reaction.

Replica-Destructive Amplification Methods

A major advantage of the replica amplification method is that because there are multiple copies of a particular array, information is not lost if a given replica is destroyed or rendered non-re-usable by a process. This allows the use of the most sensitive detection methods, regardless of their impact on the subsequent usefulness of that particular replica of the array. For example, tyramide-biotin/HRP (or other enzymatic in situ reactions) or biotin/avidin or antibody/hapten complexes (or other ligand sandwiches) may be used to effectively amplify the signal in a nucleic acid hybridization (or other bimolecular binding) experiment. These methods, however, may be considered destructive to the DNA array in that they involve interactions which are kinetically difficult to disrupt without destroying the array. Similarly, some detection processes, including sequencing by ligation and restriction and the variant methods described herein (see Examples 11 and 12), necessarily involve destruction, either chemically or enzymatically or both, of the template array. The availability of replica arrays made according to the methods disclosed herein allow the use of these methods, as they destroy only the replica, not the original or other copies. The availability of replicas of an array allows the use of direct fluorescent detection of probes hybridized to the array without loss of the array for subsequent uses. One method which this allows is the relative quantitation of mRNA by hybridization of the array with fluorescently labeled total cDNA probes. This method allows the evaluation of changes in the expression of a wide array of genes in populations of RNA isolated from cells or tissues in different growth states or following treatment with various stimuli.

Fluorescently labeled cDNA probes are prepared according to the methods described by DeRisi et al., 1997, Science 278: 680-686 and by Lockhart et al., 1996, Nature Biotechnol. 14: 1675-1680. Briefly, each total RNA (or mRNA) population is reverse transcribed from an oligo-dT primer in the presence of a nucleoside triphosphate labeled with a spectrally distinguishable fluorescent moiety. For example, one population is reverse

transcribed in the presence of Cy3-dUTP (green fluorescence signal), and another reverse transcribed in the presence of Cy5-dUTP (red fluorescence signal).

Hybridization conditions are as described by DeRisi et al. (1997, supra) and Lockhart et al. (1996, supra). Briefly, final probe volume should be 10-12 µl, at 4X SSC, and contain non-specific competitors (e.g.,, poly dA, C_0T1 DNA for a human cDNA array) as required. To this mixture is added 0.2 µl of 10% SDS and the probes are boiled for two minutes and quick chilled for ten seconds. The denatured probes are pipetted onto the array and covered with a 22mm x 22 mm cover slip. The slide bearing the array is placed in a humid hybridization chamber which is then immersed in a water bath (62°C) and incubated for 2-24 hours. Following incubation, slides are washed in solution containing 0.2X SSC, 0.1% SDS and then in 0.2X SSC without SDS. After washing, excess liquid is removed by centrifugation in a slide rack on microtiter plate carriers. The hybridized arrays are then immediately ready for scanning with a fluorescent scanning confocal microscope. Such microscopes are commercially available; details concerning design and construction of a scanner are also available on the World Wide Web at www.cmgm.stanford.edu/pbrown.

In the above example in which one population of RNA was reverse-transcription labeled with Cy3 and the other with Cy5 fluorescent dyes, the relative expression of genes represented by the features of the micro-array may be evaluated by the presence of green (Cy3, indicating the mRNA from this population hybridizes to a given feature), red (Cy5, indicating the mRNA from this population hybridizes to a given feature) or yellow (indicating that both mRNA populations used to make probes contain mRNAs which hybridize to a given feature) fluorescent signals. Alternatively, separate replicas of the same array may be hybridized separately with probes labeled with the same fluorescent dye marker but made from different populations of mRNA. For example, cDNA probes made from cells before and after treatment with a growth factor may be hybridized with separate replicas of a genomic array made from those cells. The intensity of the signal of each feature may be compared before and after growth factor treatment to yield a representation of genes induced, repressed, or whose expression is unaffected by the growth factor treatment. This method requires that the replica arrays contain one or more markers which will not vary as a means of aligning the hybridized arrays. Such a marker may be a foreign or synthetic DNA,

for example. The RNA corresponding to such a marker is spiked at equal concentration into the reverse transcription reactions used to generate labeled cDNA probes. Prior to the first hybridization with experimental cDNAs, a control hybridization using only the marker cDNA may be performed on a replica array to precisely determine the position(s) of the marker(s) within the array.

In either the simultaneous hybridization or the separate hybridization methods, the availability of additional replicas of the array allows further characterization (including but not limited to sequencing and isolation of the gene represented by the feature) of those features of the array which exhibit particular expression patterns.

EXAMPLE 3

A. Characterization of Nucleic Acid Molecules of an Array of the Invention Identification of Features of the Array

Ideally, feature identification is performed on the first array of a set produced by the methods described above; however, it is also done using any array of a set, regardless of its position in the line of production. The features are sequenced by hybridization to fluorescently labeled oligomers representing all sequences of a certain length (.e.g., all 4096 hexamers) as described for Sequencing-by-Hybridization (SBH, also called Sequencing-by-Hybridization-to-an-Oligonucleotide-Matrix, or SHOM; Drmanac et al., 1993, Science, 260(5114): 1649-52; Khrapko, et al. 1991, supra; Mugasimangalam et al., 1997, Nucleic Acids Res., 25: 800-805). The sequencing here is considerably easier than conventional SBH if the feature lengths are short (e.g., ss-25-mers rather than the greater than ds-300-mers used in SBH), if the genome sequence is known or if a preselection of features is used.

SBH involves a strategy of overlapping block reading. It is based on hybridization of DNA with the complete set of immobilized oligonucleotides of a certain length fixed in specific positions on a support. The efficiency of SBH depends on the ability to sort out effectively perfect duplexes from those that are imperfect (i.e. contain base pair mismatches). This is achieved by comparing the temperature-dependent dissociation curves of the duplexes formed by DNA and each of the immobilized oligonucleotides with standard dissociation curves for perfect oligonucleotide duplexes.

To generate a hybridization and dissociation curve, a ^{32}P-labeled DNA fragment (30,000 cpm, 30 fmoles) in 1 µl of hybridization buffer (1M NaCl; 10mM Na phosphate, pH 7.0; 0.5mM EDTA) is pipetted onto a dry plate so as to cover a dot of an immobilized oligonucleotide. Hybridization is performed for 30 minutes at 0°C. The support is rinsed with 20 ml of hybridization buffer at 0°C and then washed 10 times with the same buffer, each wash being performed for 1 minute at a temperature 5°C higher than the previous one. The remaining radioactivity is measured after each wash with a minimonitor (e.g., a Mini monitor 125; Victoreen) additionally equipped with a count integrator, through a 5mm aperture in a lead screen. The remaining radioactivity (% of input) is plotted on a logarithmic scale against wash temperature.

For hybridization with a fluorescently-labeled probe, a volume of hybridization solution sufficient to cover the array is used, containing the probe fragment at a concentration of 2 fmoles/0.01 µl. The hybridization incubated for 5.0 hour at 17°C and then washed at 0°C, also in hybridization buffer. Hybridized signal is observed and photographed with a fluorescence microscope (e.g., Leitz "Aristoplan"; input filter 510-560nm, output filter 580 nm) equipped with a photocamera. Using 250 ASA film, an exposure of approximately 3 minutes is taken.

For SBH, one suitable immobilization support is a 30 µm-thick polyacrylamide gel covalently attached to glass. Oligonucleotides to be used as probes in this procedure are chemically synthesized (e.g., by the solid-support phosphoramidite method, deprotected in ammonium hydroxide for 12 h at 55°C and purified by PAGE under denaturing conditions). Prior to use, primers are labeled either at the 5′-end with [γ-^{32}P]ATP, using T4 polynucleotide kinase, to a specific activity of about 1000 cpm/fmol, or at the 3'-end with a fluorescent label, e.g., tetramethylrhodamine (TMR), coupled to dUTP through the base by terminal transferase (Aleksandrova et al., 1990, Molek. Biologia [Moscow], 24: 1100-1108) and further purified by PAGE.

An alternative method of sequencing involves subsequent rounds of stepwise ligation and cleavage of a labeled probe to a target polynucleotide whose sequence is to be determined (Brenner, U.S. Patent No. 5,599,675). According to this method, the nucleic acid to be sequenced is prepared as a double-stranded DNA molecule with a "sticky end," in other

words, a single-stranded terminal overhang, which overhang is of a known length that is uniform among the molecules of the preparation, typically 4 to 6 bases. These molecules are then probed in order to determine the identity of a particular base present in the single-stranded region, typically the terminal base. A probe of use in this method is a double-stranded polynucleotide which (i) contains a recognition site for a nuclease, and (ii) typically has a protruding strand capable of forming a duplex with a complementary protruding strand of the target polynucleotide. In each sequencing cycle, only those probes whose protruding strands form perfectly-matched duplexes with the protruding strand of the target polynucleotide hybridize- and are then ligated to the end of the target polynucleotide. The probe molecules are divided into four populations, wherein each such population comprises one of the four possible nucleotides at the position to be determined, each labeled with a distinct fluorescent dye. The remaining positions of the duplex-forming region are occupied with randomized, unlabeled bases, so that every possible multimer the length of that region is represented; therefore, a certain percentage of probe molecules in each pool are complementary to the single-stranded region of the target polynucleotide; however, only one pool bears labeled probe molecules that will hybridize.

After removal of the unligated probe, a nuclease recognizing the probe cuts the ligated complex at a site one or more nucleotides from the ligation site along the target polynucleotide leaving an end, usually a protruding strand, capable of participating in the next cycle of ligation and cleavage. An important feature of the nuclease is that its recognition site be separate from its cleavage site. In the course of such cycles of ligation and cleavage, the terminal nucleotides of the target polynucleotide are identified. As stated above, one such category of enzyme is that of type IIs restriction enzymes, which cleave sites up to 20 base pairs remote from their recognition sites; it is contemplated that such enzymes may exist which cleave at distances of up to 30 base pairs from their recognition sites.

Ideally, it is the terminal base whose identity is being determined (in which it is the base closest to the double-stranded region of the probe which is labeled), and only this base is cleaved away by the type IIs enzyme. The cleaved probe molecules are recovered (e.g., by hybridization to a complementary sequence immobilized on a bead or other support matrix) and their fluorescent emission spectrum measured using a fluorimeter or other light-gathering

device. Note that fluorimetric analysis may be made prior to cleavage of the probe from the test molecule; however, cleavage prior to qualitative analysis of fluorescence allows the next round of sequencing to commence while determination of the identity of the first sequenced base is in progress. Detection prior to cleavage is preferred where sequencing is carried out in parallel on a plurality of sequences (either segments of a single target polynucleotide or a plurality of altogether different target polynucleotides), e.g., attached to separate magnetic beads, or other types of solid phase supports, such as the replicable arrays of the invention. Note that whenever natural protein endonucleases are employed as the nuclease, the method further includes a step of methylating the target polynucleotide at the start of a sequencing operation to prevent spurious cleavages at internal recognition sites fortuitously located in the target polynucleotide.

By this method, there is no requirement for the electrophoretic separation of closely-sized DNA fragments, for difficult-to-automate gel-based separations, or the generation of nested deletions of the target polynucleotide. In addition, detection and analysis are greatly simplified because signal-to noise ratios are much more favorable on a nucleotide-by-nucleotide basis, permitting smaller sample sizes to be employed. For fluorescent-based detection schemes, analysis is further simplified because fluorophores labeling different nucleotides may be separately detected in homogeneous solutions rather than in spatially overlapping bands.

As alluded to, the target polynucleotide may be anchored to a solid-phase support, such as a magnetic particle, polymeric microsphere, filter material, or the like, which permits the sequential application of reagents without complicated and time-consuming purification steps. The length of the target polynucleotide can vary widely; however, for convenience of preparation, lengths employed in conventional sequencing are preferred. For example, lengths in the range of a few hundred basepairs, 200-300, to 1 to 2 kilobase pairs are most often used.

Probes of use in the procedure may be labeled in a variety of ways, including the direct or indirect attachment of radioactive moieties, fluorescent moieties, colorimetric moieties, and the like. Many comprehensive reviews of methodologies for labeling DNA and constructing DNA probes provide guidance applicable to constructing probes (see

Matthews et al., 1988, Anal. Biochem., 169: 1-25; Haugland, 1992, Handbook of Fluorescent Probes and Research Chemicals, Molecular Probes, Inc., Eugene, OR; Keller and Manak, 1993, DNA Probes, 2nd Ed., Stockton Press, New York; Eckstein, ed., 1991, Oligonucleotides and Analogues: A Practical Approach, ML Press, Oxford, 1991); Wetmur, 1991, Critical Reviews in Biochemistry and Molecular Biology, 26: 227-259). Many more particular labelling methodologies are known in the art (see Connolly, 1987, Nucleic Acids Res., 15: 3131-3139; Gibson et al. 1987, Nucleic Acids Res., 15: 5455-6467; Spoat et al., 1987, Nucleic Acids Res., 15: 4837-4848; Fung et al., U.S. Pat. No. 4,757,141; Hobbs, et al., U.S. Pat. No. 5,151,507; Cruickshank, U.S. Pat. No. 5,091,519; [synthesis of functionalized oligonucleotides for attachment of reporter groups]; Jablonski et al., 1986, Nucleic Acids Res., 14: 6115-6128 [enzyme/oligonucleotide conjugates]; and Urdea et al., U.S. Pat. No. 5,124,246 [branched DNA]). The choice of attachment sites of labeling moieties does not significantly affect the ability of a given labeled probe to identify nucleotides in the target polynucleotide, provided that such labels do not interfere with the ligation and cleavage steps. In particular, dyes may be conveniently attached to the end of the probe distal to the target polynucleotide on either the 3' or 5' termini of strands making up the probe, e.g., Eckstein (cited above), Fung (cited above), and the like. In some cases, attaching labeling moieties to interior bases or inter-nucleoside linkages may be desirable.

As stated above, four sets of mixed probes are provided for addition to the target polynucleotide, where each is labeled with a distinguishable label. Typically, the probes are labeled with one or more fluorescent dyes, e.g., as disclosed by Menchen et al, U.S. Pat No. 5,188,934; Begot et al PCT application PCT/US90/ 05565. Each of four spectrally resolvable fluorescent labels may be attached, for example, by way of Aminolinker II (all available from Applied Biosystems, Inc., Foster City, Calif.); these include TAMRA (tetramethylrhodamine), FAM (fluorescein), ROX (rhodamine X), and JOE (2', 7'-dimethoxy-4',5'-dichlorofluorescein) and their attachment to oligonucleotides is described in Fung et al., U.S. Pat. No. 4,855,225.

Typically, nucleases employed in the invention are natural protein endonucleases (i) whose recognition site is separate from its cleavage site and (ii) whose cleavage results in a protruding strand on the target polynucleotide. Class IIS restriction endonucleases that may

be employed are as previously described (Szybalski et al., 1991, Gene, 100: 13-26; Roberts et al., 1993, Nucleic Acids Res., 21: 3125-3137; Livak and Brenner, U.S. Pat No. 5,093,245). Exemplary class IIs nucleases include *Alw*XI, *Bsm*AI, *Bbv*I, *Bsm*FI, *Sis*I, *Hga*I, *Bsc*AI, *Bbv*II, *Bce*fI, *Bce*85I, *Bcc*I, *Bcg*I, *Bsa*I, *Bsg*I, *Bsp*MI, *Bst*71 I, *Ear*1, *Eco*57I, *Esp*3I, *Fau*I, *Fok*I, *Gsu*I, *Hph*I, *Mbo*II, *Mme*I, *Rle*AI, *Sap*I, *Sfa*NI, *Taq*II, *Tth*1llII, *Bco*5I, *Bpu*AI, *Fin*I, *Bsr*DI, and isoschizomers thereof. Preferred nucleases include *Fok*1, *Hga*I, *Ear*I, and *Sfa*NI. Reactions are generally carried out in 50 μL volumes of manufacturer's (New England Biolabs) recommended buffers for the enzymes employed, unless otherwise indicated. Standard buffers are also described in Sambrook et al., 1989, supra.

When conventional ligases are employed, the 5' end of the probe may be phosphorylated. A 5' monophosphate can be attached to a second oligonucleotide either chemically or enzymatically with a kinase (see Sambrook et al., 1989, supra). Chemical phosphorylation is described by Horn and Urdea, 1986, Tetrahedron Lett., 27: 4705, and reagents for carrying out the disclosed protocols are commercially available (e.g., 51 Phosphate-ONTm from Clontech Laboratories; Palo Alto, Calif.).

Chemical ligation methods are well known in the art, e.g., Ferris et al., 1989, Nucleosides & Nucleotides, 8: 407-414; Shabarova et al., 1991, Nucleic Acids Res., 19: 4247-4251. Typically, ligation is carried out enzymatically using a ligase in a standard protocol. Many ligases are known and are suitable for use in the invention (Lehman, 1974, Science, 186: 790-797; Engler et al., 1982, "DNA Ligases," in Boyer, ed., The Enzymes, Vol. 15B pp. 3-30, Academic Press, New York). Preferred ligases include T4 DNA ligase, T7 DNA ligase, *E. coli* DNA ligase, Taq ligase, Pfu ligase and Tth ligase. Protocols for their use are well known, (e.g., Sambrook et al., 1989, supra; Barany, 1991, PCR Methods and Applications, 1: 5-16; Marsh et al., 1992, Strategies, 5: 73-76). Generally, ligases require that a 5' phosphate group be present for ligation to the 3' hydroxyl of an abutting strand. This is conveniently provided for at least one strand of the target polynucleotide by selecting a nuclease which leaves a 5' phosphate, e.g., *Fok*I.

Prior to nuclease cleavage steps, usually at the start of a sequencing operation, the target polynucleotide is treated to block the recognition sites and/or cleavage sites of the nuclease being employed. This prevents undesired cleavage of the target polynucleotide

because of the fortuitous occurrence of nuclease recognition sites at interior locations in the target polynucleotide. Blocking can be achieved in a variety of ways, including methylation and treatment by sequence-specific aptamers, DNA binding proteins, or oligonucleotides that form triplexes. Whenever natural protein endonucleases are employed, recognition sites can be conveniently blocked by methylating the target polynucleotide with the so-called "cognate" methylase of the nuclease being used; for most (if not all) type II bacterial restriction endonucleases, there exist cognate methylases that methylate their corresponding recognition sites. Many such methylases are known in the art (Roberts et al., 1993, supra; Nelson et al., 1993, Nucleic Acids Res., 21: 3139-3154) and are commercially available from a variety of sources, particularly New England Biolabs (Beverly, Mass.).

The method includes an optional capping step after the unligated probe is washed from the target polynucleotide. In a capping step, by analogy with polynucleotide synthesis (e.g., Andrus et al., U.S. Pat. No. 4,816,571), target polynucleotides that have not undergone ligation to a probe are rendered inert to further ligation steps in subsequent cycles. In this manner spurious signals from "out of phase" cleavages are prevented. When a nuclease leaves a 5' protruding strand on the target polynucleotides, capping is usually accomplished by exposing the unreacted target polynucleotides to a mixture of the four dideoxynucleoside triphosphates, or other chain-terminating nucleoside triphosphates, and a DNA polymerase. The DNA polymerase extends the Y strand of the unreacted target polynucleotide by one chain-terminating nucleotide, e.g., a dideoxynucleotide, thereby rendering it incapable of ligating with probe in subsequent cycles.

Alternatively, a simple method involving quantitative incremental fluorescent nucleotide addition sequencing (QIFNAS), is employed in which each end of each clonal oligonucleotide is sequenced by primer extension with a nucleic acid polymerase (e.g., Klenow or Sequenase™; U.S. Biochemicals) and one nucleotide at a time which has a traceable level of the corresponding fluorescent dNTP or rNTP, for example, 100 micromolar dCTP and 1 micromolar fluorescein-dCTP. This is done sequentially, e.g., dATP, dCTP, dGTP, dTTP, dATP and so forth until the incremental change in fluorescence is below a percentage that is adequate for useful discrimination from the cumulative total from previous cycles. The length of the sequence so determined may be extended by any of periodic

photobleaching or cleavage of the accumulated fluorescent label from nascent nucleic acid molecules or denaturing the nascent nucleic acid strands from the array and re-priming the synthesis using sequence already obtained.

After features are identified on a first array of the set, it is desirable to provide landmarks by which subsequently-produced arrays of the set are aligned with it, thereby enabling workers to locate on them features of interest. This is important, as the first array of a set produced by the method of the invention is, by nature, random, in that the nucleic acid molecules of the starting pool are not placed down in a specific or pre-ordered pattern based upon knowledge of their sequences.

Several types of markings are made according to the technology available in the art. For instance, selected features are removed by laser ablation (Matsuda and Chung 1994, ASAIO Journal, 40(3): M594-7; Jay, 1988, Proc. Natl. Acad. Sci. U.S.A., 85: 5454-5458; Kimble, 1981, Dev. Biol., 87(2): 286-300) or selectively replicated on copies of an array by laser-enhanced adhesion (Emmert-Buck et al, 1996, Science, 274(5289): 998-1001). These methods are used to eliminate nucleic acid features that interfere with adjacent features or to create a pattern that is easier for software to align.

Laser ablation is carried out as follows: A KrF excimer laser, e.g., a Hamamatsu L4500 (Hamamatsu, Japan) (pulse wavelength, 248nm; pulse width, 20ns) is used as the light source. The laser beam is converged through a laser-grade UV quartz condenser lens to yield maximum fluences of 3.08 J/cm^2 per pulse. Ablation of the matrix and underlying glass surface is achieved by this method. The depth of etching into the glass surfaces is determined using real-time scanning laser microscopy (Lasertec 1LM21W, Yokohama, Japan), and a depth profile is determined.

Selective transfer of features *via* laser-capture microdissection proceeds as follows: A flat film (100μm thick) is made by spreading a molten thermoplastic material e.g., ethylene vinyl acetate polymer (EVA; Adhesive Technologies; Hampton, NH) on a smooth silicone or polytetrafluoroethylene surface. The optically-transparent thin film is placed on top of an array of the invention, and the array/film sandwich is viewed in an inverted microscope (e.g., and Olympus Model CK2; Tokyo) at 100× magnification (10× objective). A pulsed carbon dioxide laser beam is introduced by way of a small front-surface mirror coaxial with the

condenser optical path, so as to irradiate the upper surface of the EVA film. The carbon dioxide laser (either Apollo Company model 580, Los Angeles, or California Laser Company model LS150, San Marcos, CA) provides individual energy pulses of adjustable length and power. A ZnSe lens focuses the laser beam to a target of adjustable spot size on the array. For transfer spots of 150 μm diameter, a 600-microsecond pulse delivers 25-30 mW to the film. The power is decreased or increased approximately in proportion to the diameter of the laser spot focused on the array. The absorption coefficient of the EVA film, measured by Fourier transmission, is 200 cm^{-1} at a laser wavelength of 10.6 μm. Because >90% of the laser radiation is absorbed within the thermoplastic film, little direct heating occurs. The glass plate or chip upon which the semi-solid support has been deposited provides a heat sink that confines the full-thickness transient focal melting of the thermoplastic material to the targeted region of the array. The focally-molten plastic moistens the targeted tissue. After cooling and recrystallization, the film forms a local surface bond to the targeted nucleic acid molecules that is stronger than the adhesion forces that mediate their affinity for the semi-solid support medium. The film and targeted nucleic acids are removed from the array, resulting in focal microtransfer of the targeted nucleic acids to the film surface.

If removal of molecules from the array by this method is performed for the purpose of ablation, the procedure is complete. If desired, these molecules instead are amplified and cloned out, as described in Example 4.

A method provided by the invention for the easy orientation of the nucleic acid molecules of a set of arrays relative to one another is "array templating." A homogeneous solution of an initial library of single-stranded DNA molecules is spread over a photolithographic all-10-mer ss-DNA oligomer array under conditions which allow sequences comprised by library members to become hybridized to member molecules of the array, forming an arrayed library where the coordinates are in order of sequence as defined by the array. For example, a 3'-immobilized 10-mer (upper strand), binds a 25-mer library member (lower strand) as shown below:

5'-TGCATGCTAT-3' [SEQ ID NO:14]

3'-CGATGCATTTACGTAACGTACGATA-5' [SEQ ID NO:15]

Covalent linkage of the 25-mer sequence to the support, amplification and replica printing are performed by any of the methods described above. Further characterization, if required, is carried out by SBH, fluorescent dNTP extension or any other sequencing method applicable to nucleic acid arrays, such as are known in the art. This greatly enhances the ability to identify the sequence of a sufficient number of oligomer features in the replicated array to make the array useful in subsequent applications.

Replica Sequencing with Ligation/Restriction Cycles

The sequencing by ligation and restriction method of Brenner, as described above, provides a powerful approach to the simultaneous sequencing of entire arrays of DNA molecules. The ability to replicate the entire array provides a novel approach to improving the efficiency of the sequencing method. In its standard format, the number of bases sequenced by the ligation and restriction method is limited by a background of molecules which fail to ligate or cleave properly in a given cycle. This phenomenon disturbs the synchrony of the process and limits the effective lengths which may be sequenced by this method since the interference it introduces is cumulative.

The sequencing by ligation and restriction method as disclosed by Brenner addresses this issue by the optional inclusion of a "capping" step after the unligated probe has been removed. According to that method, when the target molecules have a 5' protruding end, a mixture of dideoxynucleoside triphosphates and a DNA polymerase is added prior to the next cleavage step. This results in the addition of a single dideoxynucleotide to the 3' terminus of the recessed strand which will prevent subsequent ligation steps, effectively deleting the molecule which failed to be ligated from the target population. The effectiveness of the capping method is dependent on the completeness of the cap addition.

An improvement on the method of sequencing by ligation and cleavage involves the use of two or more distinct probes comprising different "ligation cassettes" coupled with a round of replica amplification by PCR wherein one of the primers is specific to the most recently added ligation cassette. This method will be referred to as "replica sequencing with ligation and restriction cycles." A probe of use in this method is a double-stranded polynucleotide which (i) contains a recognition site for a nuclease, (ii) typically has a

protruding strand capable of forming a duplex with a complementary protruding strand of the target polynucleotide, and (iii) which has a sequence, the "ligation cassette," such that an oligonucleotide primer complementary to one such sequence or cassette will allow amplification of the molecule to which it is ligated under the conditions used for annealing and extension within the method.

In each sequencing cycle, only those probes whose protruding strands form perfectly-matched duplexes with the protruding strand of the target polynucleotide hybridize and are then ligated to the end of the target polynucleotide. The probe molecules are divided into four populations, wherein each such population comprises one of the four possible nucleotides at the position to be determined, each labeled with a distinct fluorescent dye. The remaining positions of the duplex-forming region are occupied with randomized, unlabeled bases, so that every possible multimer the length of that region is represented; therefore, a certain percentage of probe molecules in each pool are complementary to the single-stranded region of the target polynucleotide; however, only one pool bears labeled probe molecules that will hybridize.

The individual probes comprising different ligation cassettes may have a recognition sequence for the same or different type IIs restriction endonuclease. The important factor is that the ligation cassette sequences, due to their distinct primer binding characteristics, allow amplification of only those target molecules which were successfully ligated in the previous ligation step. This also enforces the requirement for completing the cleavage step, as those target molecules which were not cleaved in the previous step will similarly not be amplified, since they will not bear the proper primer. This process enriches the proportion of each feature which has successfully completed the most recent cycle of ligation and restriction. Through the reduction in background due to improved synchrony, this method increases the number of bases which can be sequenced for features on a given array. The added steps of the replication and subsequent re-amplification of the array not only further enrich for sequences which are in synchrony, but also confers control over the size of the features, as described herein in the section entitled "Geometrical Focusing." As discussed in that section, control over the size of the features with increasing numbers of amplification or replication

cycles allows more sequence or other information to be gleaned from a given array before features begin to overlap.

After a cycle of cleavage, ligation of a first ligation cassette, and subsequent detection of the next base in the sequence, the steps one will perform in applying the replica amplification process to this method of sequencing are as follows: 1) using primers, one complementary to the common end (arbitrarily designated the 5' end, for this discussion) of the features being sequenced, and the other complementary to the most recently added ligation cassette, the features of the array are amplified and then replicated according to methods described herein above; 2) a replica is then subjected to a new cycle of cleavage, ligation of a probe comprising a distinct ligation cassette, and detection of the next base in the sequence; 3) the features of the array are amplified using the primer complementary to the common 5' end of the features and a primer complementary to the distinct ligation cassette, followed by replication of the array; and 4) the process of steps 1-3 is repeated until the sequences of the features are determined.

Within the method of replica sequencing with ligation and restriction cycles, a new probe comprising a distinct ligation cassette sequence may be used for each cycle of ligation and restriction. Alternatively, fewer different ligation cassettes than the number of cycles of ligation and restriction may be used. In other words, as few as two and as many as n (where n equals the number of cycles of ligation and restriction) different ligation cassettes may be of use according to the method. As used herein, "new" or "different" or "distinct" when referring to probes or ligation cassettes comprised by probes is meant to indicate that the sequence of each ligation cassette, or the oligonucleotide probe comprising it, is such that a primer complementary to the ligation cassette will not hybridize with any other cassette or oligonucleotide comprising a cassette under the conditions used for annealing and polymerization. Clearly, the greater the number of different ligation cassettes used, the more strictly the requirement for completion of previous cycles will be enforced. It is within the ability of one of skill in the art to determine how many different ligation cassettes are required to achieve a desired level of synchrony (with a concomitant reduction in background). As a general guideline, since the background due to incomplete cycles is cumulative, the number of ligation cassettes will vary in proportion to the desired number of

bases to be sequenced. One would, for example, expect to use a larger number of different ligation cassettes if 300 bases are to be sequenced than one would use to sequence 30 bases.

Replication of the arrays in the method of replica sequencing by ligation and restriction may be performed as often as every cycle, once every nth cycle (where n is greater than 1), or even once per whole set of cycles. Again, the frequency of replication may be determined by one skilled in the art. Considerations include, but are not limited to the physical size of the features and the overall desired number of bases to be sequenced.

The method of Jones, 1997, Biotechniques 22: 938-946 teaches the use of PCR amplification to positively select for those molecules in a population which had successfully completed the previous cycle of cleavage and ligation. Jones did not, however, teach the replication of amplified populations or the application of the method to random arrays of features. Rather, Jones taught the use of microwell plates and a robotic pipetting apparatus to perform his method. An important advantage of the incorporation of the replication step into the sequencing method is that it allows control over the size of the amplified features. While Jones mentions the eventual application of his method to the "biochip" format, no guidance is given which would allow one to overcome the inherent limitation on the size of the features in a method incorporating PCR amplification steps on a microarray. In contrast, novel methods based on the replication of arrays, such as geometrical focusing, are described herein which overcome this limitation.

Non-Replica Sequencing

Methods allowing determination of DNA sequences on an array that do not involve replica production are also preferred for some applications. For example, sequencing of transcription products (or their reverse transcripts) in situ requires that the fine resolution of the sequencing templates be preserved.

One may use the method of Jones (1997, supra) to sequence features on an array without replicating the array. Other non-electrophoretic methods which might be adapted to sequencing of microarrays include the single nucleotide addition methods of minisequencing (Canard & Sarfati, 1994, Gene 148: 1-6; Shoemaker et al., 1996, Nature Genet. 14: 450-456; Pastinen et al., 1997, Genome Res. 7: 606-614; Tully et al., 1996, Genomics 34: 107-113;

Jalanko et al., 1992, Clin. Chem. 38: 39-43; Paunio et al., 1996, Clin. Chem. 42: 1382-1390; Metzker et al., 1994, Nucl. Acids Res. 22: 4259-4267) and pyrosequencing (Uhlen & Lundeberg, U.S. Patent No. 5,534,424; Ronaghi et al., 1998, Science 281: 363-365; Ronaghi et al., 1999, Anal. Biochem. 267: 65-71).

As an alternative to minisequencing or pyrosequencing, the novel method of fluorescent in situ sequencing extension quantification (FISSEQ) may be used. FISSEQ involves the following steps: 1) a mixture of primer, buffer and polymerase are added to a microarray of single stranded DNA; 2) a single, fluorescently labeled base is added to the mixture, and will be incorporated if it is complementary to the corresponding base on the template strand; 3) unincorporated dNTP is washed away; 4) incorporated dNTP is detected by monitoring fluorescence; 5) steps 2-4 are repeated (using fresh buffer and polymerase) with each of the four dNTPs in turn; and 6) steps 2-5 are repeated in cycles until the sequence is known.

The method of sequencing nucleic acid molecules within a polyacrylamide gel matrix using the Fluorescent In Situ Sequencing Extension Quantification method and nucleotides labeled with cleavable linkers was demonstrated in the following experiments.

In order to evaluate the method, molecules of a known DNA sequence were first cast into a polyacrylamide gel matrix. The oligonucleotide sequencing primer RMGP1-R (5' - gcc cgg tct cga gcg tct gtt ta) was annealed to the oligonucleotide puc514c (Q - 5' tcggcc aacgcgcggg gagaggcggt ttgcgtatca g **taaacagac gctcgagacc gggc** (sample 1)) or to the oligonucleotide puc234t (Q - 5' cccagt cacgacgttg taaaacgacg gccagtgtcg a **taaacagac gctcgagacc gggc** (sample 2). The bolded sequences denote the sequences to which the sequencing primer anneals, and Q indicates an ACRYDITE modification.

Equal amounts of template and primer were annealed at a final concentration of 5μM in 1x EcoPol buffer (10mM Tris pH 7.5, 5mM $MgCl_2$), by heating to 95 degrees C for 1 minute, slowly cooling to 50 degrees C at a rate of 0.1 degrees per second, and holding the reaction at 50 degrees C for 5 minutes. The primer:template complex was then diluted by adding 30μl 1x Ecopol buffer and 2μl 500mM EDTA.

One microliter of each annealed oligonucleotide was added to 17μl of acrylamide gel mixture (40mM Tris pH 7.3, 25% glycerol, 1mM DTT, 6% acrylamide (5% cross-linking),

17.4 units SEQUENASE version 2.0 (United States Biochemical, USB), 15μg/ml *E. coli* single stranded binding protein (USB), 0.1mg/ml BSA). Then, 1μl of 1.66% TEMED and 1μl of 1.66% APS were added and 0.2μl of each mixture was pipetted onto bind-silane treated glass microscope slides. The slides were immediately put under an argon bed for 30 minutes to allow polymerization of the acrylamide.

The slides containing the spots of polyacrylamide containing DNA molecules to be sequenced were then washed in 40mM Tris pH 7.5, 0.01% Triton X-100 for 30 seconds, after which the slides were ready for sequencing reactions. Each slide was subjected to a number of single nucleotide extension cycles (in the nomenclature adopted for the purposes of this example, a single nucleotide extension cycle means the addition of one nucleotide, not the sequential addition of each of the four nucleotides G, A, T, and C). For each cycle, the slide was incubated in extension buffer with one nucleotide for 4 minutes at room temperature. Between cycles, the slides were washed twice for minutes each in FISSEQ wash buffer (10mM Tris pH 7.5, 250mM NaCl, 2mM EDTA, 0.01% Triton X-100), and spun briefly to dry. Slides were scanned on a GSI SCANARRAY 4000 fluorescence scanner.

In the first cycle, each slide was incubated in dATP extension mix (10mM Tris pH 7.5 50mM NaCl, 5mM $MgCl_2$, 0.1mg/ml BSA, 0.01% Triton X-100, 0.2 μM unlabeled dATP). In the next cycle each slide was incubated in the dCTP extension mix (as above, with dCTP replacing dATP). In all, Slide 1 was subjected to 5 cycles of unlabeled nucleotide addition (i.e., A, then C, then G, then T, then A), followed by 1 cycle of fluorescently labeled dCTP addition (10mM Tris pH 7.5 50mM NaCl, 5mM $MgCl_2$, 0.1mg/ml BSA, 0.01% Triton X-100, 0.2 μM unlabeled dCTP, 0.2 μM Cy3-dCTP).

Figure 1 shows a fluorescence scan of slide 1 after the cycle in which the labeled dCTP was added, above a schematic of the sequencing templates indicating the expected extension products for each template. Fluorescent label was detected in spots containing sample 1, where the sixth template nucleotide is a G, which allows the addition of the labeled C to the primer. No label was detected in spots containing sample 2, which agrees with the fact that the next template nucleotide was a T, which did not allow incorporation of the labeled C onto the primer. These data indicate that sequencing reactions in polyacrylamide

spots remain in phase after 6 additions, and that misincorportion by the polymerase is not high under these conditions.

A second slide, slide 2, was subjected to 7 cycles of unlabeled nucleotide addition (i.e., A, then C, then G, then T, then A, then C, then G), followed by 1 cycle of Cy5-dUTP addition (10mM Tris pH 7.5 50mM NaCl, 5mM MgCl2, 0.1mg/ml BSA, 0.01% Triton X-100, 0.2 μM unlabeled dTTP, 0.2 μM Cy5-dUTP). Figure 2 shows a scan of slide 2 after the Cy5-dUTP addition, and a schematic of the expected extension products. Since both nucleic acid sequencing template samples 1 and 2 encoded an A as the next base to be added to the primer, no signal is detected in spots containing either sample template. This confirms that the sequences were maintained in phase through 6 additions, and further indicates a lack of misincorporation by the polymerase under these conditions.

Slide 3 was subjected to 9 cycles of unlabeled nucleotide addition (A, then C, then G, then, T, then A, then C, then G, then T, then A) followed by 1 cycle of Cy3-dCTP addition. The fluorescence scan of slide 3 is shown in Figure 3. Fluorescently labeled C was correctly added to the primer on sample 1, but was not added to the primer on sample 2.

Finally, slide 4 was subjected to 11 cycles of unlabeled nucleotide addition (A, then C, then G, then T, then A, then C, then G, then T, then A, then C, then G), followed by 1 cycle of Cy5-dUTP addition. The fluorescence scan of slide 4 after the labeled dUTP cycle (Figure 4) shows that dUTP was correctly added to the primer on sample 2.

The experiments shown in Figures 1-4 establish that the fluorescent in situ sequencing extension quantification method permits sequencing of at least twelve nucleotides on a template contained within a polyacrylamide gel. There was no indication of misincorporation by the polymerase under these conditions. Further, as shown by the similar detection of signal in each of 5 spots containing a given nucleic acid sequencing template in a given cycle, the sequencing reactions remained in phase for at least twelve nucleotide additions. There is no reason to believe further nucleotide additions would not be possible using these methods. In addition, any of the methods described herein below to further extend the sequence read length of the FISSEQ method may be used.

It is recognized that polymerases used for sequencing become inefficient for further extension when 100% of bases added to a primer are non-native (i.e., fluorescently labeled).

Therefore, the efficiency of FISSEQ may be further improved by employing a mixture of native and fluorescently labeled dNTP. The mixture allows incorporation of labeled bases at each position without requiring 100% adjacent non-native bases. Also, a photobleaching step after each set of one or more cycles may be incorporated to allow the computational background subtraction to act on a smaller number, with corresponding lower Poisson shot noise.

As an alternative to photobleaching or computational subtraction of accumulating fluorescence, cleavable linkages between the fluorophore and the nucleotide may be employed to permit removal of the fluorophore after incorporation and detection, thereby setting the sequence up for additional labeled base addition and detection. As used herein, the term "cleavable linkage" refers to a chemical moiety that joins a fluorophore to a nucleotide, and that can be cleaved to remove the fluorophore from the nucleotide when desired, essentially without altering the nucleotide or the nucleic acid molecule it is attached to. Cleavage may be accomplished, for example, by acid or base treatment, or by oxidation or reduction of the linkage, or by light treatment (photobleaching), depending upon the nature of the linkage. Examples of cleavable linkages are described by Shimkus et al., 1985, Proc. Natl. Acad. Sci. USA 82: 2593-2597; Soukup et al., 1995, Bioconjug. Chem. 6: 135-138; Shimkus et al., 1986, DNA 5: 247-255; and Herman and Fenn, 1990, Meth. Enzymol. 184: 584-588, all of which are incorporated herein by reference.

As one example of a cleavable linkage, a disulfide linkage may be reduced using thiol compound reducing agents such as dithiothreitol. Fluorophores are available with a sulfhydryl (SH) group available for conjugation (e.g.,, Cyanine 5 or Cyanine 3 fluorophores with SH groups; New England Nuclear - DuPont), as are nucleotides with a reactive aryl amino group (e.g.,, dCTP). A reactive pyridyldithiol will react with a sulfhydryl group to give a sulfhydryl bond that is cleavable with reducing agents such as dithiothreitol. An NHS-ester heterobifunctional crosslinker (Pierce) is used to link a deoxynucleotide comprising a reactive aryl amino group to a pyridyldithiol group, which is in turn reactive with the SH on a fluorophore, to yield a disulfide bonded, cleavable nucleotide-fluorophore complex useful in the methods of the invention (see, for example, Figure 5). Alternatively, a cis-glycol linkage between a nucleotide and a fluorophore can be cleaved by periodate. These are

examples of standard components of cleavable cross-linkers used for protein chemistry or for polyacrylamide gels. In this embodiment, cleavage of the fluorophore could be done as often as every cycle, or less frequently, such as every other, every third, or every fifth or more cycles.

A modified embodiment of FISSEQ that allows longer effective reads involves extension for a fixed number of cycles with mixtures of three native (unlabeled) dNTPs interspersed with pulses of wash, up to a desired length. Following this, one begins cycles of adding one partially labeled (i.e., mixture of labeled and unlabeled) dNTP at a time. The triple dNTP cycles allow positioning of the polymerase a fixed distance from the primer and would use alternating sets of triphosphates (e.g.,, ACG, CGT, ACG, ...) chosen and concentration optimized to reduce false incorporation and failure to incorporate (Hillebrand et al., 1984, Nucl. Acids. Res. 12: 3155-3171). This allows three times longer reads plus any advantage possibly conferred by having fewer potential misincorporation steps. It is contemplated that if the misincorporation rate (n-1 and extensible n+1 products) can be as low as 10^{-4}, then read lengths longer than current electrophoresis-based methods are possible.

Another modification using the triple dNTP cycles is aimed at reducing the background caused by mismatch incorporation. If, for example, G:T mismatch pairing is a major source of misincorporation (Keohavong et al., 1993, PCR Meth. Appl. 2: 288-292), one should always include A with G, since the more stable A:T interaction will be favored over the less stable G:T interaction. For example, one may alternate triple mix 1 (dATP, dCTP, dGTP) with triple mix 2 (dCTP, dGTP, dTTP).

A more conservative version of FISSEQ which can allow determination of longer stretches of sequences at a time requires replicas of the array, and will be referred to as replica-FISSEQ. Replica arrays for this method may be made by the replica amplification methods described herein, or by a microarray spotting method using a microarray robot. By spotting the same DNA templates in known positions on the slide, the same effect can be obtained as with the replica-amplified features. In this embodiment, 30 identical arrays are made using the microarray robot. Stepping through 1 to 30 additions with native (unlabeled) dNTPs sets up the final base to be assessed for each array element (e.g.,, slide 1 gets zero native base additions, slide 2 gets one native base addition, etc.). The final base is assessed

by the sequential addition of each fluorescent dNTP as is normally done in minisequencing. Pyrosequencing data (Ronaghi et al., 1998, Science 281: 363) has shown that the polymerase extension reactions stay accurately in phase through at least 30 cycles of dNTP addition using natural nucleotides and Klenow exo- polymerase. To read out N bases with the single slide method described above requires 4N cycles of nucleotide addition and washing. The N-slide (triple dNTP, 4 cycles per slide) method (using N replicas), requires 2N(N-1)/3 cycles. The actual read lengths will be more than N bases (1.4N on average due to runs of identical bases). The same number of scans are required for the two methods.

Several other modifications to the basic method of FISSEQ are contemplated. For example, a loop may be incorporated into the primer to help reduce mispriming events (Ronaghi et al., 1998, Biotechniques 25: 876-878, 880-882, and 884). A particularly useful loop structure, described by Hirao et al. (1994, Nucl. Acids Res. 22: 576-582) as "extraordinarily stable," would have the advantage of having a relatively short stem, lowering the stability of the complementary strand hairpin, the result being that the asymmetric PCR for the strand that we want will extend to the correct end more efficiently.

Another modification would address the difficulty, encountered in many methods, of sequencing past long repeating stretches. If it is known that a given array contains many such sequences, one may include a defined regimen (for example, halfway through the whole sequence) of deoxy- and dideoxynucleotides to reduce out-of-phase templates. That is, if one knows he or she is sequencing through a repeat of, for example, AC dinucleotides, one may reduce the number of out-of-phase molecules by following a dATP addition with a ddATP addition. Only those molecules which failed to incorporate the deoxy- form of the nucleotide will be available to incorporate the dideoxy- form, leading to chain termination and reduction of that source of background. Clearly, similar regimens may be devised for repeats involving more than two nucleotides. It should be noted that the strategy is not limited to repeats and may be used to extend read length in any situation where most of the sequences in the array have a block of sequence part of the way through the target sequence which is known. For example, in an array of targets, most having the unique sequence ACGTA at the same distance from the primer, one may reduce the number of out-of-phase molecules by

following a dATP addition with a ddATP, ddGTP, and ddTTP addition, then dCTP followed by ddATP, ddCTP, and ddTTP addition.

Gel Sequencing of Amplified Array Features Using Dye Terminators

In addition to the methods of sequencing by hybridization and sequencing by ligation and restriction, it is possible to sequence amplified features of arrays using fluorescently labeled dideoxynucleoside triphosphates ("dye terminators") using the Sanger ("dideoxy") sequencing method (Sanger et al., 1975, J. Mol. Biol., 94:441) and a micro gel system. In this embodiment, the array of amplified features is created in a linear arrangement along one edge of a very thin slab gel or at the edge of a microfabricated array of capillaries. DNA molecules of the pool to be sequenced are prepared in any of the same ways as for the random array spot format described above, such that each molecule in the pool has a known sequence or sequences at one or both ends which may serve as primer binding sites. The DNA is applied to the slide as in the random array format, except that it is restricted to a thin line, rather than a circular spot. Alternatively, the DNA may be derived as a replica of a line within a standard 2D array, or may be derived as a replica of a line from a metaphase chromosome spread.

Features of the deposited linear array are then amplified using any of the methods described above for amplification of spot arrays. This amplification may be linear or exponential, thermocycled or isothermal. Isothermal amplification methods include the Phi29 rolling circle amplification method (Lizardi et al., 1998, Nature Genetics 19: 225-232), reverse transcriptase / T4 DNA polymerase / Klenow / T7 RNA polymerase linear amplification (Phillips and Eberwine, 1996, Methods 10: 283-288) and a T7 DNA polymerase / thioredoxin / ssb system (Tabor and Richardson, January 1999 Department of Energy Human Genome Program Abstract No. 15; PCT/US00/00580).

The amplified DNA template may be replicated using the methods described above. This template, which is immobilized either covalently, by entanglement, or by steric hindrance of the gel (or other semi-solid) is then reacted with dye terminators in the presence of the other necessary components of the dideoxy sequencing method (i.e., primer, dNTPs, buffer and polymerase). It is well known in the art that a number of polymerases may be

used for dideoxy-sequencing, including but not limited to Klenow polymerase, Sequenase™ or Taq polymerase. A major advantage of dye terminators over fluorescently labeled primers ("dye primers") is that the use of dye terminators requires only one reaction containing four distinguishably labeled terminators, whereas the use of dye primers requires four separate reactions which would require four identical amplified features and software alignment of the post-size-separation pattern. It should be noted that dye terminators also exist for RNA polymerase sequencing (Sasaki et al., 1998, Proc. Natl. Acad. Sci. USA 95: 3455-3460). It should also be noted that if the termination reactions have been performed with the use of primers, then a rare-cutting endonuclease may be used to produce a desired end for the sequencing ladder.

A miniature gel system appropriate for the gel sequencing of linear feature arrays has been described by Stein et al., 1998, Nucl. Acids. Res. 26: 452-455. In this system, small, ultrathin polyacrylamide gels are cast, eight or more at a time, on standard microscope slides. The gels may be stored, ready to use, for approximately two weeks. They are run horizontally in a standard mini-agarose gel apparatus, with typical run times of 6 to 8 minutes. Stein et al. describe a novel sample loading system which permits volumes as low as 0.1 μl to be analyzed. The band resolution compares favorably with that of large-format sequencing gels. Within the context of the sequencing of linear arrays according to the invention, the sample loading is accomplished by performing the termination reactions within, or at the very edge of the gel, rather than by mechanical means.

Since the terminated reaction products remain bound to the template, the reaction may be cleaned of dNTPs, primers and salts by diffusion, flow and/or electrophoresis. The termination products are then denatured and electrophoresed perpendicular to the line of amplified features in a thin slab or capillary format. An important aspect of this method is that the order of the amplified features is preserved throughout the process. Thus, if the line of features comes from a chromosome or large cloned or uncloned DNA fragment, the long range order is preserved and greatly aids in the assembly of complex genomic regions even in the presence of long repeats. Similarly, if the lines of features are derived as replicas of lines from the standard 2D arrays, the sequence identity of each spot in that line may be determined. Similar replicas of additional lines from the 2D spot may be used to determine

the identity of each spot or feature of the 2D array. In addition to the clear advantages regarding the spatial organization of the features, this method has the additional advantage of actually using more of the sequencing reaction than other methods. That is, all of the reaction products are electrophoresed, rather than just a portion of it, meaning there is less waste of reagents. Further, the immobilization of the features allows the use of a common pool of reagents to sequence many features simultaneously. Thus, the method is more economical on a per sequence basis.

Fluorescent *in Situ* Sequencing Extension Quantification with Cleavable Linkers

The method of sequencing nucleic acid molecules within a polyacrylamide gel matrix using the Fluorescent In Situ Sequencing Extension Quantification method and nucleotides labeled with cleavable linkers was demonstrated in the following experiments.

In order to evaluate the method, molecules of a known DNA sequence were first cast into a polyacrylamide gel matrix. The oligonucleotide sequencing primer RMGP1-R (5' - gcc cgg tct cga gcg tct gtt ta) was annealed to the oligonucleotide puc514c (Q - 5' tcggcc aacgcgcggg gagaggcggt ttgcgtatca g **taaacagac gctcgagacc gggc** (sample 1)) or to the oligonucleotide puc234t (Q - 5' cccagt cacgacgttg taaaacgacg gccagtgtcg a **taaacagac gctcgagacc gggc** (sample 2). The bolded sequences denote the sequences to which the sequencing primer anneals, and Q indicates an ACRYDITE modification.

Equal amounts of template and primer were annealed at a final concentration of 5μM in 1x EcoPol buffer (10mM Tris pH 7.5, 5mM MgCl2), by heating to 95 degrees C for 1 minute, slowly cooling to 50 degrees C at a rate of 0.1 degrees per second, and holding the reaction at 50 degrees C for 5 minutes. The primer:template complex was then diluted by adding 30μl 1x Ecopol buffer and 2μl 500mM EDTA.

One microliter of each annealed oligonucleotide was added to 17μl of acrylamide gel mixture (40mM Tris pH 7.3, 25% glycerol, 1mM DTT, 6% acrylamide (5% cross-linking), 17.4 units SEQUENASE version 2.0 (United States Biochemical, USB), 15μg/ml *E. coli* single stranded binding protein (USB), 0.1mg/ml BSA). Then, 1μl of 1.66% TEMED and 1μl of 1.66% APS were added and 0.2μl of each mixture was pipetted onto bind-silane

treated glass microscope slides. The slides were immediately put under an argon bed for 30 minutes to allow polymerization of the acrylamide.

The slides containing the spots of polyacrylamide containing DNA molecules to be sequenced were then washed in 40mM Tris pH 7.5, 0.01% Triton X-100 for 30 seconds, after which they were ready for the incorporation of labeled nucleotides. For this experiment, dCTP labeled with the fluorophore Cy5 with either a non-cleavable linkage (referred to herein as Cy5-dCTP) or with a disulfide-containing cleavable linkage (referred to herein as Cy5-SS-dCTP) was used. The acrylamide spots containing known DNA to be sequenced were incubated in 30 µl of Cy-5 dCTP extension mix (10mM Tris pH 7.5 50mM NaCl, 5mM MgCl2, 0.1mg/ml BSA, 0.01% Triton X-100, 0.1 µM unlabeled dCTP, 0.2 µM Cy5-dCTP) or in Cy-5-SS-dCTP extension mix (10mM Tris pH 7.5 50mM NaCl, 5mM $MgCl_2$, 0.1mg/ml BSA, 0.01% Triton X-100, 0.1µM unlabeled dCTP, 0.2 µM Cy5-SS-dCTP) for 4 minutes at room temperature. The slides were washed twice, for 5 minutes each in FISSEQ wash buffer (10mM Tris pH 7.5, 250mM NaCl, 2mM EDTA, 0.01% Triton X-100), spun briefly to dry and scanned on a Scanarray 4000 confocal scanner (GSI Luminomics). The settings were as follows: Focus = 2060, Laser = 80%, PMT = 80% resolution = 30 microns.

Cleavage of the cleavable disulfide linkages was performed by incubation with the reducing agent dithiothreitol (DTT). The slides were incubated overnight in FISSEQ wash buffer supplemented with 5 mM DTT, washed twice for 5 minutes each in wash buffer, spun briefly to dry and scanned as before. Figure 6 shows the results of this experiment. Sample 1 incorporated both the cleavable and the non-cleavable fluorescently labeled nucleotide (see "Before DTT Wash" panels), while sample 2 did not, as was expected since only sample 1 had a G as the next template nucleotide. DTT wash (bottom panels) removed the fluorescent signal from the samples extended with the Cy5-SS-dCTP sample, but not from the samples extended with the non-cleavably linked fluorophore, demonstrating that the cleavable linkages could be cleaved, or chemically bleached, from the Cy5-SS-dCTP-extended samples with reducing agent, but not from the Cy5-dCTP-extended samples. One of skill in the art would fully expect similar cleavable linkages to nucleotides other than dCTP (for example, dATP, dGTP, TTP or even ribonucleotides or further modified nucleotides) to function in a similar manner.

Enhancing the Performance of Nucleic Acid Sequencing in Polyacrylamide-immobilized Arrays

Polyacrylamide-immobilized nucleic acid arrays and replicas thereof, made as described herein above or through other methodologies, are useful as platforms for simultaneously sequencing the large number of different DNA molecules comprising the array. In particular, the FISSEQ methods described herein above, in all variations, are useful approaches to sequencing DNAs in polyacrylamide-immobilized arrays. There are a number of parameters of the polyacrylamide gels and sequencing conditions that may be modified to enhance the performance of the FISSEQ method (also referred to as ISAS, or "In Situ Amplification and Sequencing) when performed on polyacrylamide-immobilized arrays.

One parameter that can be modified is the pore size of the gel. Larger pore size allows the polymerase(s) used for thermal cycling, sequencing, or both, to diffuse more freely and access the primed template. In the sequencing reactions, increased pore size increases the efficiency of base addition so that rapid "dephasing" or loss of synchrony of the template strands is prevented. Depending on the crosslinker and total acrylamide concentration, standard acrylamide pore sizes are generally about 5 to about 20 nanometers. For example, in gels with 5% total acrylamide and 4% bis-acrylamide cross linker, the pore size is about 5 nm. There are several methods known for creating so-called "macroporous" polyacrylamide gels, with pores of about 100 nm to about 600 nm in diameter. As used herein, the term "macroporous polyacrylamide gel" refers to a polyacrylamide gel with pore size of about 25 to 600 nm in diameter, with a preferred range of about 100 to about 600 nm.

First, polyethylene glycol (PEG) may be added to the gel. See for example, Righetti et al., 1992, Electrophoresis 13: 587-595, incorporated herein by reference, which describes gel polymerization in the presence of "laterally aggregating agents" such as PEG to increase pore size. A preferred preparation uses 6% acrylamide, 1.5% cross-linker (e.g.,, bis-acrylamide), with 2.5% PEG (10 kDa polymer size). The total acrylamide may be varied over a range from about 3% to about 12%, and the cross-linker may vary from about 1% to about 30%. All percentages are weight per volume. In these formulations, the PEG may be varied from 0% to about 25%, with the polymer size of the PEG molecules varying from about 1 kDa to about 20 kDa. Generally, the longer the PEG chain length, the lower the

percentage of PEG needed to increase the pore size. The inclusion of PEG in the polyacrylamide gel results in pores up to approximately 100 times the size of those achievable using acrylamide alone.

Alternatively, N,N'-diallyltartardiamide (DATD) may be used as the cross linking agent. See for example, Spath and Koblet, 1979, Anal. Biochem. 93: 275-285, incorporated herein by reference, which compares DATD-cross-linked gels to Bis-acrylamide cross-linked gels.

As another alternative, it is known that polymerization at low temperatures results in larger pore sizes in polyacrylamide gels. Standard practice for polyacrylamide gel polymerization is to perform the reaction at room temperature. However, polymerization at 4°C produces a gel with larger pore sizes compared to a gel of the same composition polymerized at room temperature. Generally, lower or reduced temperatures for gel polymerization include a range from about 0°C to about 15°C, with a temperature of about 2°C to about 4°C being preferred. Polymerization at 4°C in a 5% total acrylamide, 4% bis-acrylamide gel, for example, results in a pore size of about 30 nm, compared to pores of about 5-20 nm when the same gel is polymerized at room temperature (i.e., about 21°C).

As another alternative, increasing the percentage of cross-linker (e.g.,, bis-acrylamide) in the acrylamide monomer solution is also known to result in a gel with larger pore size relative to gels formed with lower percentages of cross-linker (see Righetti et al., 1981, J. Biochem. Biophys. Meth. 4: 347-363, which is incorporated herein by reference). As noted above, cross-linker may be varied from about 1% to about 30%, with higher percentages yielding greater pore sizes.

In addition to gel pore size, another parameter that can be manipulated to enhance the efficiency of sequencing reactions in polyacrylamide array gels is the amount of secondary structure of the template DNAs. For example, single-stranded binding protein (SSBP) may be added to the sequencing reaction in order to reduce the amount of secondary structure of the template molecules. Reduced secondary structure reduces pausing by the polymerase that can contribute to dephasing of the reactions on an array. Generally, *E. coli* SSBP (U.S. Biochemical) is added to the sequencing reactions at concentrations ranging from about 1 μM to about 5 μM.

Salt conditions are also important in the amount of template secondary structure and may be varied to enhance sequencing efficiency on polyacrylamide-immobilized arrays. Generally, intramolecular interactions contributing to secondary structure are reduced as salt concentration is decreased. It is acknowledged that different polymerases useful in the methods of the invention can have different sensitivities to and requirements for salt concentrations. One of skill in the art is readily able to determine the effect of decreasing salt concentration on a given polymerase with respect to sequencing fidelity and efficiency. Useful salt concentrations generally range from about 2 to about 10 mM $MgCl_2$ and about 0 to 100 mM NaCl. Exemplary salt conditions for sequencing include the following: for Klenow fragment of E. coli DNA polymerase, 10 mM $MgCl_2$, without any NaCl; for Sequenase, 50 mM NaCl and 5 mM $MgCl_2$; for Bst polymerase, 50 mM NaCl and 5 mM $MgCl_2$.

Preferred conditions for sequencing polyacrylamide-immobilized DNA array features include 50 mM NaCl, and 5 μM SSBP, at room temperature using 0.5 μM Sequenase.

The temperature of the reaction may also be varied to enhance the efficiency of DNA sequencing reactions within the gel, as this also affects the secondary structure of the template molecules. Generally, the secondary structure is reduced as the temperature of the reaction is increased. It is helpful, therefore, to use a thermostable polymerase such as Bst polymerase (New England Biolabs) or Thermosequenase (Amersham).

When using higher temperatures for sequencing reactions it is helpful or sometimes even necessary to increase the length of the sequencing primer or the G+C content of the primer/primer binding sequences in order to determine the maximum temperature (T_m) at which primer annealing is maintained while reducing intramolecular template secondary structure. One of skill in the art may calculate the T_m for a given oligonucleotide primer at a given salt concentration. As an example, however, for primers greater than 10 bases in a 50 mM salt solution (standard PCR conditions), T_m may be estimated using the formula T_m= 59.9 + 41[%G+C (decimal value)] - [675/primer length].

B. Characterization of Polypeptide Members of an Array of the Invention

The present invention provides for protein arrays as well as nucleic acid arrays, in which protein arrays are either derived through expression of arrayed nucleic acids or alternatively, though binding of proteins to arrayed nucleic acids. Arrayed proteins can be characterized by any of a number of means known in the art. For example, proteins can be identified through antibody-based methodologies, mass spectroscopy, labeled ligand:receptor binding interactions, as well as assays of enzymatic activities of proteins that are present on an array.

EXAMPLE 4

Isolation of a Feature of the Array

Isolation of Nucleic Acid Molecules (Method 1)/ Heterologous Arrays

As described above in Example 1, sets of arrays are, if desired, produced according to the invention such that they incorporate oligonucleotide sequences bearing restriction sites linked to the ends of each feature. This provides a method for creating spatially-unique arrays of primer pairs for *in situ* amplification, in which each feature has a distinct set of primer pairs. One or both of the universal primers comprises a restriction endonuclease recognition site, such as a type IIS sequence (e.g., as *Eco*57I or *Mme*I which will cut up to 20 bp away). Treatment of the whole double-stranded array with the corresponding enzyme(s) followed by melting and washing away the non-immobilized strand creates the desired primer pairs with well-defined 3' ends. Alternatively, a double-strand-specific 3' exonuclease treatment of the double-stranded array is employed, but the resulting single-stranded 3' ends may vary in exact endpoint. The 3' end of the primers are used for in situ amplification, for example of variant sequences in diagnostics. This method, by which arrays of unique primer pairs are produced efficiently, provides an advance over the method of Adams and Kron (1997, supra), in which each single pair of primers is manually constructed and placed. Cloning of a given feature from an array of such a set is performed as follows:

MmeI is a restriction endonuclease having the property of cleaving at a site remote from its recognition site, TCCGAC. Heterogeneous pools of primers are constructed that comprise (from 5' to 3') a sequence shared by all members of the pool, the MmeI recognition

site, and a variable region. The variable region may comprise either a fully-randomized sequence (e.g., all possible hexamers) or a selected pool of sequences (e.g., variations on a particular protein-binding, or other, functional sequence motif). If the variable sequence is random, the length of the randomized sequence determines the sequence complexity of the pool. For example, randomization of a hexameric sequence at the 3' ends of the primers results in a pool comprising 4,096 distinct sequence combinations. Examples of two such mixed populations of oligonucleotides (in this case, 32-mers) are primer pools 1s and 2s, below:

primer 1 (a pool of 4096 32-mers):

5' gcagcagtacgactagcataTCCGACnnnnnn 3' [SEQ ID NO:16]

primer 2 (a pool of 4096 32-mers):

5' cgatagcagtagcatgcaggTCCGACnnnnnn 3' [SEQ ID NO:17]

A nucleic acid preparation is amplified, using primer 1 to randomly prime synthesis of sequences present therein. The starting nucleic acid molecules are cDNA or genomic DNA, either of which may comprise molecules that are substantially whole or that are into smaller pieces. Many DNA cleavage methods are well known in the art. Mechanical cleavage is achieved by several methods, including sonication, repeated passage through a hypodermic needle, boiling or repeated rounds of rapid freezing and thawing. Chemical cleavage is achieved by means which include, but are not limited to, acid or base hydrolysis, or cleavage by base-specific cleaving substances, such as are used in DNA sequencing (Maxam and Gilbert, 1977, Proc. Natl. Acad. Sci. U.S.A., 74: 560-564). Alternatively, enzymatic cleavage that is site-specific, such as is mediated by restriction endonucleases, or more general, such as is mediated by exo- and endonucleases e.g., ExoIII, mung bean nuclease, DNAase I or, under specific buffer conditions, DNA polymerases (such as T4), which chew back or internally cleave DNA in a proofreading capacity, is performed. If the starting nucleic acid molecules (which may, additionally, comprise RNA) are fragmented rather than whole (whether closed circular or chromosomal), so as to have free ends to which

a second sequence may be attached by means other than primed synthesis, the *Mme*I recognition sites may be linked to the starting molecules using DNA ligase, RNA ligase or terminal deoxynucleotide transferase. Reaction conditions for these enzymes are as recommended by the manufacturer (e.g., New England Biolabs; Beverly, MA or Boehringer Mannheim Biochemicals, Indianapolis, IN). If employed, PCR is performed using template DNA (at least 1 fg; more usefully, 1-1,000 ng) and at least 25 pmol of oligonucleotide primers; an upper limit on primer concentration is set by aggregation at about 10 μg/ml. A typical reaction mixture includes: 2μl of DNA, 25 pmol of oligonucleotide primer, 2.5 μl of 10× PCR buffer 1 (Perkin-Elmer, Foster City, CA), 0.4 μl of 1.25 μM dNTP, 0.15 μl (or 2.5 units) of Taq DNA polymerase (Perkin Elmer, Foster City, CA) and deionized water to a total volume of 25 μl. Mineral oil is overlaid and the PCR is performed using a programmable thermal cycler. The length and temperature of each step of a PCR cycle, as well as the number of cycles, is adjusted in accordance to the stringency requirements in effect. Initial denaturation of the template molecules normally occurs at between 92°C and 99°C for 4 minutes, followed by 20-40 cycles consisting of denaturation (94-99°C for 15 seconds to 1 minute), annealing (temperature determined as discussed below, 1-2 minutes), and extension (72°C for 1 minute). Final extension is generally for 4 minutes at 72°C, and may be followed by an indefinite (0-24 hour) step at 4°C.

Annealing temperature and timing are determined both by the efficiency with which a primer is expected to anneal to a template and the degree of mismatch that is to be tolerated. In attempting to amplify a mixed population of molecules, the potential loss of molecules having target sequences with low melting temperatures under stringent (high-temperature) annealing conditions against the promiscuous annealing of primers to sequences other than their target sequence is weighed. The ability to judge the limits of tolerance for feature loss vs. the inclusion of artifactual amplification products is within the knowledge of one of skill in the art. An annealing temperature of between 30°C and 65°C is used. An example of one primer out of the pool of 4096 primer 1, one primer (primer 1ex) is shown below, as is a DNA sequence from the preparation with which primer 1ex has high 3' end complementarity at a random position. The priming site is underlined on either nucleic acid molecule.

primer 1ex [SEQ ID NO:19; bases 1 - 32]: 5'-gcagcagtacgactagcataTCCGAC ctgcgt-3'

genomic DNA [SEQ ID NO:18]: 3'-tttcgacgcacatcgcgtgcatggccccatgcatcagg
ctgacgaccgtcgtacgtctactcggct-5'

After priming, polymerase extension of primer 1ex on the template results in:
[SEQ ID NO:19] 5'-gcagcagtacgactagcataTCCGACctgcgtgtagcgcacgtaccggggtacgtagtcc
gactgctggcagcatgcagatgagccga-3'

Out of the pool of 4096 primer 2, one primer with high 3' end complementarity to a random position in the extended primer 1ex DNA is selected by a polymerase for priming (priming site in bold):

[SEQ ID NO:19] 5'-gcagcagtacgactagcataTCCGACctgcgtgtagcgcacgtaccggggtacgtagtcc
ga**ctgctg**gcagcatgcagatgagccga 3'

primer 2ex [SEQ ID NO:20; bases 1-32]: 3'-**gacgac**CAGCCTggacgtacgatgacgatagc-5'

After priming and synthesis, the resulting second strand is:
[SEQ ID NO:20] 3'-cgtcgtcatgctgatcgtatAGGCTGgacgcacatcgcgtgcatggccccatgcatcagg
ct**gacgac**CAGCCTggacgtacgatgacgatagc-5'

Primer 3, shown below, is a 26-mer that is identical to the constant region of primer 1ex:
[SEQ ID NO:19; nucleotides 1-26] 5'-gcagcagtacgactagcataTCCGAC-3'
It is immobilized by a 5' acrylyl group to a polyacrylamide layer on a glass slide.

Primer 4, below, is a 26-mer that is complementary to the constant region of primer 2ex:
[SEQ ID NO:20; nucleotides 1-26] 5'-cgatagcagtagcatgcaggTCCGAC-3'
It is optionally immobilized to the polyacrylamide layer by a 5' acrylyl group.

The pool of amplified molecules derived from the sequential priming of the original nucleic acid preparation with mixed primers 1 and 2, including the product of 1ex/2ex priming and extension, are hybridized to immobilized primers 3 and 4. *In situ* PCR is performed as described above, resulting in the production of a first random, immobilized array of nucleic acid molecules according to the invention. This array is replicated by the methods described in Example 2 in order to create a plurality of such arrays according to the invention.

After *in situ* PCR using primers 3 and 4:

5'-gcagcagtacgactagcataTCCGACctgcgtgtagcgcacgtaccggggtacgtagt

3'-cgtcgtcatgctgatcgtatAGGCTGgacgcacatcgcgtgcatggccccatgcatca

ccgactgctgGTCGGAcctgcatgctactgctatcg-3' [SEQ ID NO:21]

ggctgacgacCAGCCTggacgtacgatgacgatagc-5' [SEQ ID NO:20]

After cutting with MmeI and removal of the non-immobilized strands:

[SEQ ID NO:21; bases 1-46] 5'-gcagcagtacgactagcataTCCGACctgcgtgtagcgcacgtacc-3'
(primer 1-based, clone-specific oligonucleotide)

[SEQ ID NO:20; bases 1-46] 3'-ccatgcatcaggctgacgacCAGCCTggacgtacgatgacgatagc-5'
(primer 2-based, clone-specific oligonucleotide)

The resulting random arrays of oligonucleotide primers representing the nucleic acid sequences of the original preparation are useful in several ways. Any particular feature, such as the above pair of primers, is used selectively to amplify the intervening sequence (in this case two central bp of the original 42 bp cloned segment are captured for each use of the chip or a replica) from a second nucleic acid sample. This is performed in solution or *in situ*, as described above, following feature identification on the array, using free, synthetic primers. If desired, allele-specific primer extension or subsequent hybridization is performed.

Importantly, this technique provides a means of obtaining corresponding, or homologous, nucleic acid arrays from a second cell line, tissue, organism or species according to the invention. The ability to compare corresponding genetic sequences derived from different sources is useful in many experimental and clinical situations. By "corresponding genetic sequences," one means the nucleic acid content of different tissues of a single organism or tissue-culture cell lines. Such sequences are compared in order to study the cell-type specificity of gene regulation or mRNA processing or to observe chromosomal rearrangements that might arise in one tissue rather than another. Alternatively, the term refers to nucleic acid samples drawn from different individuals, in which case a given gene or its regulation is compared between or among samples. Such a comparison is of use in linkage studies designed to determine the genetic basis of disease, in forensic techniques and in population genetic studies. Lastly, it refers to the characterization and comparison of a particular nucleic acid sequence in a first organism and its homologues in one or more other organisms that are separated evolutionarily from it by varying lengths of time in order to highlight important (therefore, conserved) sequences, estimate the rate of evolution and/or establish phylogenetic relationships among species. The invention provides a method of generating a plurality of immobilized nucleic acid arrays, wherein each array of the plurality contains copies of nucleic acid molecules from a different tissue, individual organism or species of organism.

Alternatively, a first array of oligonucleotide primers with sequences unique to members of a given nucleic acid preparation is prepared by means other than the primed synthesis described above. To do this, a nucleic acid sample is obtained from a first tissue, cell line, individual or species and cloned into a plasmid or other replicable vector which comprises, on either side of the cloning site, a type IIS enzyme recognition site sufficiently close to the junction between vector and insert that cleavage with the type IIS enzyme(s) recognizing either site occurs within the insert sequences, at least 6 to 10, preferably 10 to 20, base pairs away from the junction site. It is contemplated that type IIS restriction endonuclease activity may even occur at a distance of up to 30 pairs from the junction site. The nucleic acid molecules are cleaved from the vector using restriction enzymes that cut outside of both the primer and oligonucleotide sequences, and are then immobilized on a

semi-solid support according to the invention by any of the methods described above in which covalent linkage of molecules to the support occurs at their 5' termini, but does not occur at internal bases. Cleavage with the type IIS enzyme (such as MmeI) to yield the immobilized, sequence-specific oligonucleotides is performed as described above in this Example.

As mentioned above, it is not necessary to immobilize primer 4 on the support. If primer 4 is left free, the *in situ* PCR products yield the upper (primer 1 derived) strand upon denaturation:

[SEQ ID NO:21] 5'-gcagcagtacgactagcataTCCGACctgcgtgtagcgcacgtaccggggtacgtagtcc gactgctgGTCGGAcctgcatgctactgctatcg-3'.

This sequence is available for hybridization to fluorescently-labeled DNA or RNA for mRNA quantitation or genotyping.

Isolation of Nucleic Acid Molecules (Method 2)

As described above, laser-capture microdissection is performed in order to help orient a worker using the arrays of a set of arrays produced according to the invention, or to remove undesirable features from them. Alternatively, this procedure is employed to facilitate the cloning of selected features of the array that are of interest. The transfer of the nucleic acid molecules of a given feature or group of features from the array to a thin film of EVA or another heat-sensitive adhesive substance is performed as described above. Following those steps, the molecules are amplified and cloned as follows:

The transfer film and adherent cells are immediately resuspended in 40 µl of 10 mM Tris·HCl (pH 8.0), 1 mM EDTA and 1% Tween-20, and incubated overnight at 37°C in a test tube, e.g., a polypropylene microcentrifuge tube. The mixture is then boiled for 10 minutes. The tubes are briefly spun (1000 rpm, 1 min.) to remove the film, and 0.5 µl of the supernatant is used for PCR. Typically, the sheets of transfer film initially applied to the array are small circular disks (diameter 0.5 cm). For more efficient elution of the after LCM transfer, the disk is placed into a well in a 96-well microliter plate containing 40 µl of extraction buffer. Oligonucleotide primers specific for the sequence of interest may be

designed and prepared by any of the methods described above. PCR is then performed according to standard methods, as described in the above examples.

EXAMPLE 5

Size-Sorted Genomic Arrays

As mentioned above, it is possible to prepare a support matrix in which are embedded whole, even living, cells. Such protocols have been developed for various purposes, such as encapsulated, implantable cell-based drug-delivery vehicles, and the delivery to an electophoretic matrix of very large, unsheared DNA molecules, as required for pulsed-field gel electrophoresis (Schwartz and Cantor, 1984, Cell, 37: 67-75). The arrays of the invention are constructed using as the starting material genomic DNA from a cell of an organism that has been embedded in an electrophoretic matrix and lysed *in situ*, such that intact nucleic acid molecules are released into the support matrix environment. If an array based upon copies of large molecules is made, such as is of use in a fashion similar to the chromosomal element ordering arrays described below in Example 7, then a low-percentage agarose gel is used as a support. Following lysis (Schwartz and Cantor, 1984, supra), the resulting large molecules may be size-sorted electrophoretically prior to *in situ* PCR amplification and linkage to the support, both as described above. If it is desired to preserve the array on a support other than agarose, which may be difficult to handle if the gel is large, the array is transferred *via* electroblotting onto a second support, such as a nylon or nitrocellulose membrane prior to linkage.

If it is not considered essential to preserve the associations between members of genetic linkage groups (at the coarsest level of resolution, chromosomes), nucleic acid molecules are cleaved, mechanically, chemically or enzymatically, prior to electrophoresis. A more even distribution of nucleic acid over the support results, and physical separation of individual elements from one another is improved. In such a case, a polyacrylamide, rather than agarose, gel matrix is used as a support. The arrays produced by this method do, to a certain extent, resemble sequencing gels; cleavage of an electrophoresed array, e.g., with a second restriction enzyme or combination thereof, followed by electrophoresis in a second dimension improves resolution of individual nucleic acid sequences from one another.

Such an array is constructed to any desired size. It is now feasible to scan large gels (for example, 40 cm in length) at high resolution. In addition, advances in gel technology now permit sequencing to be performed on gels a mere 4 cm long, one tenth the usual length, which demonstrates that a small gel is also useful according to the invention.

EXAMPLE 6

RNA Localization Arrays

The methods described in Example 7, below, are applied with equal success to the generation of an array that provides a two-dimensional representation of the spatial distribution of the RNA molecules of a cell. This method is applied to 'squashed' cellular material, prepared as *per* the chromosomal spreads described below in Example 7; alternatively, sectioned tissue samples affixed to glass surfaces are used. Either paraffin-, plastic- or frozen (Serrano et al., 1989, Dev. Biol. 132: 410-418) sections are used in the latter case.

Tissue samples are fixed using conventional reagents; formalin, 4% paraformaldehyde in an isotonic buffer, formaldehyde (each of which confers a measure of RNAase resistance to the nucleic acid molecules of the sample) or a multi-component fixative, such as FAAG (85 % ethanol, 4% formaldehyde, 5% acetic acid, 1% EM grade glutaraldehyde) is adequate for this procedure. Note that water used in the preparation of any aqueous components of solutions to which the tissue is exposed until it is embedded is RNAase-free, i.e. treated with 0.1% diethylprocarbonate (DEPC) at room temperature overnight and subsequently autoclaved for 1.5 to 2 hours. Tissue is fixed at 4°C, either on a sample roller or a rocking platform, for 12 to 48 hours in order to allow fixative to reach the center of the sample. Prior to embedding, samples are purged of fixative and dehydrated; this is accomplished through a series of two- to ten-minute washes in increasingly high concentrations of ethanol, beginning at 60%- and ending with two washes in 95%- and another two in 100% ethanol, followed two ten-minute washes in xylene. Samples are embedded in any of a variety of sectioning supports, e.g., paraffin, plastic polymers or a mixed paraffin/polymer medium (e.g., Paraplast®Plus Tissue Embedding Medium, supplied by Oxford Labware). For example, fixed, dehydrated tissue is transferred from the second

xylene wash to paraffin or a paraffin/polymer resin in the liquid-phase at about 58°C, then replace three to six times over a period of approximately three hours to dilute out residual xylene, followed by overnight incubation at 58°C under a vacuum, in order to optimize infiltration of the embedding medium in to the tissue. The next day, following several more changes of medium at 20 minute to one hour intervals, also at 58°C, the tissue sample is positioned in a sectioning mold, the mold is surrounded by ice water and the medium is allowed to harden. Sections of 6μm thickness are taken and affixed to 'subbed' slides, which are those coated with a proteinaceous substrate material, usually bovine serum albumin (BSA), to promote adhesion. Other methods of fixation and embedding are also applicable for use according to the methods of the invention; examples of these are found in Humason, G.L., 1979, Animal Tissue Techniques, 4th ed. (W.H. Freeman & Co., San Francisco), as is frozen sectioning.

Following preparation of either squashed or sectioned tissue, the RNA molecules of the sample are reverse-transcribed *in situ*. In order to contain the reaction on the slide, tissue sections are placed on a slide thermal cycler (e.g., Tempcycler II; COY Corp., Grass Lake, MI) with heating blocks designed to accommodate glass microscope slides. Stainless steel or glass (Bellco Glass Inc.; Vineland, NJ) tissue culture cloning rings approximately 0.8 cm (inner diameter) X 1.0 cm in height are placed on top of the tissue section. Clear nail polish is used to seal the bottom of the ring to the tissue section, forming a vessel for the reverse transcription and subsequent localized *in situ* amplification (LISA) reaction (Tsongalis et al., 1994, supra).

Reverse transcription is carried out using reverse transcriptase, (e.g., avian myoblastosis virus reverse transcriptase, AMV-RT; Life Technologies/Gibco-BRL or Moloney Murine Leukemia Virus reverse transcriptase, M-MLV-RT, New England Biolabs, Beverly, MA) under the manufacturer's recommended reaction conditions. For example, the tissue sample is rehydrated in the reverse transcription reaction mix, minus enzyme, which contains 50 mM Tris·HCl (pH 8.3), 8 mM $MgCl_2$, 10 mM dithiothreitol, 1.0 mM each dATP, dTTP, dCTP and dGTP and 0.4 mM oligo-dT (12- to 18-mers). The tissue sample is, optionally, rehydrated in RNAase-free TE (10 mM Tris·HCl, pH 8.3 and 1 mM EDTA), then drained thoroughly prior to addition of the reaction buffer. To denature the RNA molecules,

which may have formed some double-stranded secondary structures, and to facilitate primer annealing, the slide is heated to 65°C for 1 minute, after which it is cooled rapidly to 37°C. After 2 minutes, 500 units of M-MLV-RT are added the mixture, bringing the total reaction volume to 100µl. The reaction is incubated at 37°C for one hour, with the reaction vessel covered by a microscope cover slip to prevent evaporation.

Following reverse transcription, reagents are pipetted out of the containment ring structure, which is rinsed thoroughly with TE buffer in preparation for amplification of the resulting cDNA molecules.

The amplification reaction is performed in a total volume of 25 µl, which consists of 75 ng of both the forward and reverse primers (for example the mixed primer pools 1 and 2 of Example 4) and 0.6 U of Taq polymerase in a reaction solution containing, per liter: 200 nmol of each deoxynucleotide triphosphate, 1.5 mmol of $MgCl_2$, 67 mmol of Tris·HCl (pH 8.8), 10 mmol of 2-mercaptoethanol, 16.6 mmol of ammonium sulfate, 6.7 µmol of EDTA, and 10 µmol of digoxigenin-11-dUTP. The reaction mixture is added to the center of the cloning ring, and layered over with mineral oil to prevent evaporation before slides are placed back onto the slide thermal cycler. DNA is denatured *in situ* at 94°C for 2 min prior to amplification. LISA is accomplished by using 20 cycles, each consisting of a 1-minute primer annealing step (55°C), a 1.5-min extension step (72°C), and a 1-min denaturation step (94°C). These amplification cycle profiles differ from those used in tube amplification to preserve optimal tissue morphology, hence the distribution of reverse transcripts and the products of their amplification on the slide.

Following amplification, the oil layer and reaction mix are removed from the tissue sample, which is then rinsed with xylene. The containment ring is removed with acetone, and the tissue containing the amplified cDNA is rehydrated by washing three times in approximately 0.5 ml of a buffer containing 100 mM Tris-Cl (pH 7.5) and 150 mM NaCl. The immobilized nucleic acid array of the invention is then formed by contacting the amplified nucleic acid molecules with a semi-solid support and covalently crosslinking them to it, by any of the methods described above.

Features are identified using SBH, also as described above, and correlated with the positions of mRNA molecules in the cell.

Protein Characterization Arrays

In addition to localizing RNA molecules in the cells, the present invention provides means of characterizing proteins in cells through use of a protein array. Nucleic acid arrays including, but not limited to, the types described herein can be expressed to form a protein array of the invention, or alternatively, can have proteins bound to the nucleic acids to achieve a protein array. For example, nucleic acid arrays can be expressed to generate protein arrays as described above in Example 1. These protein arrays can be used, for example, to determine which proteins might co-localize with particular cells or tissues.

EXAMPLE 7

Ordered Chromosomal Arrays According to the Invention

Direct *in situ* single-copy (DISC)-PCR is a method that uses two primers that define unique sequences for on-slide PCR directly on metaphase chromosomes (Troyer et al., 1994a, Mammalian Genome, 5: 112-114; summarized by Troyer et al., 1997, Methods Mol. Biol., Vol. 71: PRINS and *In Situ* PCR Protocols, J.R. Godsen, ed., Humana Press, Inc., Totowa, NJ, pp. 71-76). It thus allows exponential accumulation of PCR product at specific sites, and so may be adapted for use according to the invention.

The DISC-PCR procedure has been used to localize sequences as short as 100-300bp to mammalian chromosomes (Troyer et al., 1994a, supra; Troyer et al., 1994b, Cytogenet. Cell Genetics, 67(3), 199-204; Troyer et al., 1995, Anim. Biotechnology, 6(1): 51-58; and Xie et al., 1995, Mammalian Genome 6: 139-141). It is particularly suited for physically assigning sequence tagged sites (STSs), such as microsatellites (Litt and Luty, 1989, Am. J. Hum. Genet , 44: 397-401; Weber and May, 1989, Am. J. Hum. Genet 44, 338-396), many of which cannot be assigned by *in situ* hybridization because they have been isolated from small-insert libraries for rapid sequencing. It can also be utilized to map expressed sequence tags (ESTs) physically (Troyer, 1994a, supra; Schmutz et al., 1996, Cytogenet. Cell Genetics, 72: 37-39). DISC-PCR obviates the necessity for an investigator to have a cloned gene in hand, since all that is necessary is to have enough sequence information to synthesize PCR primers. By the methods of the invention, target-specific primers need not even be utilized; all that is required is a mixed pool of primers whose members have at one end a 'universal'

sequence, suitable for manipulations such as restriction endonuclease cleavage or hybridization to oligonucleotide molecules immobilized on- or added to a semi-solid support and, at the other end, an assortment of random sequences (for example, every possible hexamer) which will prime *in situ* amplification of the chromosome. As described above, the primers may include terminal crosslinking groups with which they may be attached to the semi-solid support of the array following transfer; alternatively, they may lack such an element, and be immobilized to the support either through ultraviolet crosslinking or through hybridization to complementary, immobilized primers and subsequent primer extension, such that the newly-synthesized strand becomes permanently bound to the array. The DISC-PCR procedure is summarized briefly as follows:

Metaphase chromosomes anchored to glass slides are prepared by standard techniques (Halnan, 1989, in Cytogenetics of Animals, C.R.E. Halnan, ed., CAB International, Wallingford, U.K., pp. 451-456;), using slides that have been pre-rinsed in ethanol and dried using lint-free gauze. Slides bearing chromosome spreads are washed in phosphate-buffered saline (PBS; 8.0 g NaCl, 1.3 g Na_2HPO_4 and 4 g NaH_2PO_4 dissolved in deionized water, adjusted to a volume of 1 liter and pH of 7.4) for 10 min and dehydrated through an ethanol series (70-, 80-, 95-, and 100%). Note that in some cases, overnight fixation of chromosomes in neutral-buffered formalin followed by digestion for 15 minutes with pepsinogen (2 mg/ml; Sigma) improves amplification efficiency.

Alternatively, chromosomes can be anchored to slides at chromosomal regions that are typically a few kb or so apart through triplex structures that are formed by any of several means. These means include formation of triplex strands at specific sequences (which are typically rich in runs of pyrimidines) of double-stranded chromosomes, recA-mediated invasion of double-stranded DNA by oligonucleotides, or peptide nucleic acids (PNA) under low salt conditions which encourage PNA invasion of the double-stranded chromosomal. In addition to triplex structures, chromosomal DNA can alternatively be anchored by other sequence-specific DNA binding molecules. Oligonucleotides, PNAs or other sequence-specific DNA binding molecules used to anchor chromosomes can be deposited at a specific, high resolution (nanometer scale) on slides by conventional methods such as inkjet printing, photolithography, and stamping. The anchoring molecules can be deposited at locations

separated by distances that are less than the maximum physical length of 0.34 microns per kilobasepair expected for double-stranded DNA. In this manner, conventional micron resolution methods can be used to align and engineer nanometer scale periodic or aperiodic arrays.

For each slide of anchored chromosomes, the following solution is prepared in a microfuge tube: 200 μM each dATP, dCTP, dGTP and dTTP; all deoxynucleotides are maintained as frozen, buffered 10 mM stock solutions or in dry form, and may be obtained either in dry or in solution from numerous suppliers (e.g., Perkin Elmer, Norwalk, CT; Sigma, St. Louis, MO; Pharmacia, Uppsala, Sweden). The reaction mixture for each slide includes 1.5 μM each primer (from 20 μM stocks), 2.0 μL 10X *Taq* polymerase buffer (100mM Tris-HCl, pH 8.3, 500 mM KCl, 15mM $MgCl_2$, 0.1% BSA; Perkin Elmer), 2.5 units Ampli*Taq* polymerase (Perkin Elmer) and deionzed H_20 to a final volume of 20 μl. Note that the commercially supplied *Taq* polymerase buffer is normally adequate; however, adjustments may be made as needed in [$MgCl_2$] or pH, in which case an optimization kit, such as the Opti-Primer PCR Kit (Stratagene; La Jolla, CA) may be used. The above reaction mixture is pipetted onto the metaphase chromosomes and covered with a 22 x 50 mm coverslip, the perimeter of which is then sealed with clear nail polish. All air bubbles, even the smallest, are removed prior to sealing, as they expand when heated, and will inhibit the reaction. A particularly preferred polish is Hard As Nails (Sally Hansen); this nail enamel has been found to be resistant to leakage, which, if it occurred, would also compromise the integrity of the reaction conditions and inhibit amplification of the chromosomal DNA sequences. One heavy coat is sufficient. After the polish has been allowed to dry at room temperature, the edges of the slide are covered with silicone grease (Dow Corning Corporation, Midland, MI). Slides are processed in a suitable thermal cycler (i.e. one designed for on-slide PCR, such as the BioOven III; Biotherm Corp., Fairfax, VA) using the following profile:

a. 94°C for 3 min.

b. Annealing temperature of primers for I min.

c. 72°C for 1 min.

d. 92°C for 1 min.

e. Cycle to step b 24 more times (25 cycles total).

f. Final extension step of 3-5 min.

After thermal cycling is complete, silicone grease is removed with a tissue, and the slide is immersed in 100% ethanol. Using a sharp razor blade, the nail polish is cut through and the edge of the coverslip is lifted gently and removed. It is critical that the slide never be allowed to dry from this point on, although excess buffer is blotted gently off of the slide edge. The slide is immersed quickly in 4X SSC and excess nail polish is scraped from the edges of the slide prior to subsequent use.

The slide is contacted immediately with a semi-solid support in order to transfer to it the amplified nucleic acid molecules; alternatively, that the slide is first equilibrated in a liquid medium that is isotonic with- or, ideally, identical to that which permeates (i.e. is present in the pores of-) the semi-solid support matrix. From that point on, the array is handled comparably with those prepared according to the methods presented in Example 1. Feature identification, also as described above, permits determination of the approximate positions of genetic elements along the length of the template chromosome. In preparations in which chromosomes are linearly extended (stretched), the accuracy of gene ordering is enhanced. This is particularly useful in instances in which such information is not known, either through classical or molecular genetic studies, even in the extreme case of a chromosome that is entirely uncharacterized. By this method, comparative studies of homologous chromosomes between species of interest are performed, even if no previous genetic mapping has been performed on either. The information so gained is valuable in terms of gauging the evolutionary relationships between species, in that both large and small chromosomal rearrangements are revealed. The genetic basis of phenotypic differences between different individuals of a single species, e.g., human subjects, is also investigated by this method. When template chromosomes are condensed (coiled), more information is gained regarding the *in vivo* spatial relationships among genetic elements. This may have implications in terms of cell-type specific gene transcriptional activity, upon which comparison of arrays generated from samples comprising condensed chromosomes drawn from cells of different tissues of the same organism may shed light.

While the methods by which histological samples are prepared, PCR is performed and the first copy of the chromosomal array is generated are time-consuming, multiple copies of the array are produced easily according to the invention, as described above in Example 1 and elsewhere. The ability of the invention to reproduce what would, otherwise, be a unique array provides a valuable tool by which scientists have the power to work in parallel- or perform analyses of different types upon comparable samples. In addition, it allows for the generation of still more copies of the array for distribution to any number of other workers who may desire to confirm or extend any data set derived from such an array at any time.

A variation on this use of the present invention is chromosome templating. DNA (e.g., that of a whole chromosome) is stretched out and fixed on a surface (Zimmermann and Cox, 1994, Nucleic Acids Res., 22(3): 492-497). Segments of such immobilized DNA are made single-stranded by exonucleases, chemical denaturants (e.g., formamide) and/or heat. The single stranded regions are hybridized to the variable portions of an array of single-stranded DNA molecules each bearing regions of randomized sequence, thereby forming an array where the coordinates of features correspond to their order on a linear extended chromosome. Alternatively, a less extended structure, which replicates the folded or partially-unfolded state of various nucleic acid compartments in a cell, is made by using a condensed (coiled), rather than stretched, chromosome.

EXAMPLE 8

Excluded Volume Protecting Groups

The density of features of the arrays is limited in that they must be sufficiently separated to avoid contamination of adjacent features during repeated rounds of amplification and replication. This is achieved using dilute concentrations of nucleic acid pools, but results in density limited by the Poisson distribution to a maximum of 37% occupancy of available appropriately spaced sites. In order to increase the density of features while maintaining the spacing necessary to avoid cross contamination, the following approach may be taken.

An activity which can bind the nucleic acid molecules of the pool is positioned in spots on the surface of the array support to create a capture array. The spots of the capture array are arranged such that they are separated by a distance greater than the size of the spots

(this is typically near the resolution of the intended detection and imaging devices, or approximately 3 microns). The size of the spots is set to be less than the diameter of the excluded volume of the nucleic acid polymer to be captured (for example, approximately one micron for 50 kb lambda DNA in 10 mM NaCl; please see Rybenkov et al., 1993, Proc. Natl. Acad. Sci. U.S.A. 90: 5307-5311, Zimmerman & Trach, 1991, J. Mol. Biol. 222: 599-620, and Sobel & Harpst, 1991, Biopolymers 31: 1559-1564, incorporated herein by reference, for methods of predicting excluded volumes of nucleic acids.

The "nucleic acid capture activity" of the array may be a hydrophilic compound, a compound which reacts covalently with the nucleic acid polymers of the pool, an oligonucleotide complementary to a sequence shared by all members of a pool (e.g.,, an oligonucleotide complementary to the 12 bp cohesive ends of a phage λ library, or oligonucleotide(s) complementary to one or both ends of a PCR-generated library containing large inserts and 6 to 50 bp of one strand exposed at one or both ends) or some other capture ligand including but not limited to proteins, peptides, intercalators, biotin, avidin, antibodies or fragments of antibodies or the like.

An ordered array of nucleic acid capture ligand spots may be made using a commercially-available micro-array synthesizer, modified inkjet printer (Castellino, 1997, supra), or the methods disclosed by Fodor et al. (U.S. Patent No. 5,510,270), Lockhart et al. (U.S. Patent No. 5,556,752) and Chetverin and Kramer (WO 93/17126). Alternatively, details on the design, construction and use of a micro-array synthesizer are available on the World Wide Web at www.cmgm.stanford.edu/pbrown.

An excess of nucleic acid or DNA is then applied to the surface of the microfabricated capture array. Each spot has multiple chances to bind a free nucleic acid molecule. However, once a spot has bound a nucleic acid molecule, it is protected from binding other molecules, i.e., the excluded volume of the bound DNA protects the spot from binding more than one molecule from the pool. Thus, saturation binding, or a situation very close to it, may be achieved while retaining the optimal spacing for subsequent amplification and replication.

The array resulting from this process may be amplified in situ and replicated according to methods described herein. Alternatively, or in addition, the array may be

treated in a way which decreases the excluded volume of the captured group so that additional rounds of excluded volume protecting group (EVPG) addition may be performed. Arrays produced in this manner not only increase the efficiency of the array beyond that normally allowed by the Poisson distribution, but also can be of predetermined geometry and/or aligned with other microfabricated features. In addition, such arrays allow complicated highly parallel enzymatic or chemical syntheses to be performed on large DNA arrays.

EXAMPLE 9

Applications of Protein Arrays

Applications of the protein arrays made above include quantitation of protein expression by methods analogous to mRNA expression or protein binding (as described in Church and Bulyk, co-pending application U.S.S.N. 09/132,368). The immobilized molecules can include proteins, peptides, and antibodies as direct translation products or after covalent and non-covalent modification, including but not limited to phosphorylation, glycosylation, proteases, chaperones, detergents, heat or solvent denaturaturation, metals, ions, organic, inorganic, and organometallic compounds. The mobile phase can contain similar proteins and peptides modified or not in mixtures or not, typically labeled and detected by biotin, chemiluminescence, electrochemical, isotopic, mass, fluorescence, colorimetric, and/or enzyme-linked assays.

Other applications would include fabrication of smart materials, biosensors, and biocomputers requiring precise specification of protein regions in the nanometer to micron scale. This could be done using a stretched, anchored chromosome version of the original polony amplification protocols, in this case using artificial, synthetic or semi-synthetic long (multi-protein) DNA molecules. The spacing of protein genes and/or their post-synthetic binding sites along the (artificial) chromosome would provide anchor points for three-dimension scaffolds, alignments, enzymatic activities, and two or more state macromolecular switches (the states being conformations, covalent modifications non-covalent binding site occupancy).

EXAMPLE 10

Production of Novel Properties Through Polony Interactions

Protein interactions and/or DNA-recombination that occurs at the interfaces between (different) polonies can be exploited to produce useful, novel properties. For example, two or more adjacent polonies can exchange material forming gradients or interfacial planes with distinct properties useful either for a) screening for optimal concentrations and/or optimal compositions or b) manufacturing such gradients or interfacial planes.